(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 November 2001 (15.11.2001)

PCT

(10) International Publication Number
WO 01/85791 A1

(51) International Patent Classification[7]: C07K 14/705, 16/28, C12N 15/12

(21) International Application Number: PCT/US01/15332

(22) International Filing Date: 11 May 2001 (11.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/203,217 11 May 2000 (11.05.2000) US
60/205,945 18 May 2000 (18.05.2000) US

(71) Applicant (*for all designated States except US*): **LIFESPAN BIOSCIENCES, INC.** [US/US]; 700 Blanchard Street, Seattle, WA 98121 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **BROWN, Joseph, P.** [GB/US]; 411 West Prospect Street, Seattle, WA 98119 (US). **MILLER, Margaret** [US/US]; 7155 S.E. Maker Street, Mercer Island, WA 98040 (US). **BURMER, Glenna** [US/US]; 7516 55th Place, NE, Seattle, WA 98115 (US). **FABRE-SUVER, Christine** [FR/US]; 1212 NE 63rd Street, Seattle, WA 98115 (US). **PRITCHARD, David** [US/US]; 3429 Burke Avenue, N., Seattle, WA 98103 (US).

(74) Agents: **HINSCH, Matthew, E.** et al.; Townsend and Townsend and Crew LLP, Two Embarcadero Center, Eighth Floor, San Francisco, CA 94111-3834 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/85791 A1

(54) Title: NUCLEIC ACID SEQUENCES FOR NOVEL GPCRS

(57) **Abstract:** The present invention is directed to new galanin receptors that are useful for treating and diagnosing a number of diseases and disorders, including, but not limited to, Alzheimer's disease, learning and memory disorders, hormonal problems, fat metabolism disorders, feeding disorders, pain perception disorders, diabetes, depression, etc. The present invention also provides methods for identifying modulators of galanin signaling. Such modulators are useful for treating the above-listed and other diseases and disorders.

NUCLEIC ACID SEQUENCES FOR NOVEL GPCRs

BACKGROUND OF THE INVENTION

Many physiologically important events are mediated by the binding of guanine nucleotide-binding regulatory proteins (G proteins) to G protein-coupled receptors (GPCRs). These events include vasodilation, stimulation or decrease in heart rate, bronchodilation, stimulation of endocrine secretions and enhancement of gut peristalsis, development, mitogenesis, cell proliferation and oncogenesis.

Guanine nucleotide-binding proteins are a family of proteins that transduce signals from numerous cell surface receptors to downstream intracellular effector molecules. G proteins are typically heterotrimeric proteins consisting of a guanyl-nucleotide binding alpha subunit, a beta and a gamma subunits, the latter two being tightly associated under physiological conditions (for a review, *see, e.g.*, Conklin *et al.*, *Cell* 73:631-641 (1993)). Each subunit is encoded by a separate gene. G proteins commonly cycle between two forms, depending on whether GDP or GTP is bound to the alpha subunit. Upon binding of a ligand to a G protein-coupled receptor, the GDP molecule bound to the alpha subunit is exchanged for a GTP molecule resulting in the dissociation of the α subunit from the β and γ subunits. The free alpha subunit and the beta-gamma complex are capable of transmitting a signal to downstream elements of a variety of signal transduction pathways, for example by binding to and activating adenyl cyclase. This fundamental scheme of events forms the basis for a multiplicity of different cell signaling phenomena.

The different members of the G protein coupled receptors super-family share a number of functional and structural characteristics. In particular, as described above, GPCRs have the ability to stimulate the exchange of bound GDP for GTP on associated G proteins alpha subunits in response to agonist binding. Structurally, GPCRs typically contain seven hydrophobic transmembrane segments that are suggested to be transmembrane helices of 20-30 amino acids connected by extracellular or cytoplasmic loops (*see, e.g.*, Kobilka *et al.*, *Science* 240:1310 (1988); Maggio *et al.*, *FEBS Lett.* 319:195 (1993); Maggio *et al.*, *Proc. Natl. Acad. Sci USA* 90:3103 (1993); Ridge *et al.*, *Proc. Natl. Sci USA* 91:3204 (1995); Schonenberg *et al.*, *J. Biol. Chem.* 270:18000 (1995); Huang *et al.*, *J. Biol. Chem.* 256:3802 (1981); Popot *et al.*, *J. Mol. Biol.* 198:655

(1987); Kahn and Engelman, *Biochemistry* 31:6144 (1992); Schoneberg *et al., EMBO J.* 15:1283 (1996); Wong *et al., J. Biol. Chem.* 265:6219 (1990); Monnot *et al., J. Biol. Chem.* 271:1507 (1996); Gudermann *et al., Annu. Rev. Neurosci.* 20:399 (1997); Osuga *et al., J. Biol. Chem.* 272:25006 (1997); Lefkowitz *et al., J. Biol. Chem.* 263:4993-4996 (1988); Panayotou and Waterfield, *Curr. Opinion Cell Biol.* 1:167-176 (1989); and G Protein-Coupled Receptor Database, http://www.gcrdb.uthscsa.edu). In addition to G proteins, many enzymes, such as, for example, adenylate cyclase, cGMP phosphodiesterase and phospholipase C, can act as effectors for GPCRs' signal transduction (*see, e.g.*, Kinnamon & Margolskee, *Curr. Opin. Neurobiol.* 6:506-513 (1996)).

A large variety of molecules have been shown to be ligands for GPCRs. Identified ligands include, for example, purines, nucleotides and melatonin (*e.g.*, adenosine, cAMP, NTPs, *etc.*), biogenic amines (*e.g.*, adrenaline, dopamine, histamine, acetylcholine, noradrenaline, serotonin, *etc.*), peptides (*e.g.*, angiotensin, calcitonin, chemokine, Corticotropin Releasing Factor, galanin, Growth Hormone Releasing Hormone, Gastric Inhibitory Peptride, Glucagon, Neuropeptide Y, Neurotensin, Opoiod, Thrombin, Secretin, Somatostatin, Thyrotropin Releasing Hormone, Vasopressin, Vasoactive Intestinal Peptide, *etc.*), lipids and lipid-based compounds (*e.g.*, cannabinoids, Platelet Activating Factor, *etc.*), excitatory amino acids and ions (*e.g.*, glutamate, calcium, GABA, *etc.*), toxins, *etc.* In addition, there are many "orphan" G protein-coupled receptors (*e.g.*, some olfactory G protein-coupled receptors) for which ligands have not been identified.

G protein-coupled receptors thus play a central role in transducing numerous signals and regulating cellular metabolism. Accordingly, GPCRs have been implicated in a large number of diseases, such as, Alzheimer's disease, rheumatoid arthritis, osteoarthritis, osteoporosis, amyotrophic lateral sclerosis, multiple sclerosis and atherosclerosis, asthma, depression, epilepsy, schizophrenia, Parkinson's disease, a number of sarcomas (*e.g.*, chondrosarcoma, Ewing's sarcoma, osteosarcoma, *etc.*) and carcinomas (*e.g.*, basal cell carcinoma, breast carcinoma, embryonal carcinoma, ovarian carcinoma, renal cell carcinoma, lung adenocarcinoma, lung small cell carcinoma, pancreatic carcinoma, prostate carcinoma, transitional carcinoma of the bladder, squamous cell carcinoma, thyroid carcinoma, *etc.*), psoriasis, cardiomyopathy, Crohn's disease, Duchenne muscular dystrophy, glioblastoma multiform, Hodgkin's disease,

lymphoma, macular degeneration, malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, seminoma, tuberculosis, tonsil, ulcerative colitis, *etc.*

While many GPCRs have been identified, many more remain to be discovered. In addition, the specific GPCRs involved in the different biological processes, and in particular diseases, are not known.

Galanin is a widely distributed 28 amino acid peptide hormone which has been shown to regulate a variety of biological processes, including, for example, hormone release, neurotransmitter release, nociception, feeding behavior, cognitive function and reproductive behavior.

Galanin signaling has been shown to modulate the release of a variety of neurotransmitters, including, but not limited to, acetylcholine, norepinephrine, serotonin and dopamine (*see, e.g.*, Bartfai *Crit. Rev. Neurobiol.* 7:229 (1993)). Cumulative evidence suggests that galanin acts as an inhibitory cosecreted peptide. Galanin has been postulated to impair secretion of neurotransmitters by acting at the pre-synaptic autoreceptors as well as at the post-synaptic action site of these neurotransmitters. In particular, galanin inhibits acetylcholine release into the ventral hippocampus. Galanin may thus impair memory and learning by inhibiting the cholinergic function.

Galanin is to date the only neurotransmitter that has been shown to be upregulated in Alzheimer's disease. In addition, a variety of experiments, including the central injection of galanin and the generation of transgenic mice, have shown that the overexpression and/or oversecretion of galanin impairs performance of memory and learning tasks. These results indicate that the hypertrophy of galanin pathways contributes to the cognitive deficits in Alzheimer's disease.

Galanin has further been shown to inhibit the release of vasopressin and insulin, while it stimulates the release of growth hormone, prolactin and luteinizing hormone. Galanin has been shown to play a role in the control of fat metabolism, and body adiposity, which may be mediated by its effect on insulin. Galanin inhibits insulin secretion and, conversely, insulin injection inhibits central galanin expression. Galanin acts within the medial preoptic area and paraventricular nucleus to modulate fat intake and fat metabolism, but the specific subtype of galanin receptors involved in this function are not known. Galanin also acts within the supraoptic nucleus and paraventricular nucleus to modulate fluid balance. In addition, galanin regulates feeding behavior.

Galanin may exert neurotrophic and/or neuroprotective actions within the central nervous system. Treatment of rats with galanin has been shown to reduce

behavioral impairments following brain injury. Galanin gene expression is upregulated in injured neurons and this may contribute to cell survival. Despite the substantial loss of cells within the locus ceruleus, the percentage of noradrenergic neurons that coexpress galanin mRNA is increased in Alzheimer's disease supporting the idea that galanin may exert a neuroprotective effect.

Galanin is co-localized with gonadotropin-releasing hormone (GnRH) in the medial preoptic region of several species. The pattern of coexpression exhibits sexual dimorphism in rats. In both rats and monkeys, gonadal hormones regulate galanin expression in GnRH cells. Galanin, acting within the anterior pituitary, plays a role in the regulation of luteinizing hormone release. Galanin facilitates sex behavior via actions within the medial preoptic regions.

Under normal conditions, galanin has potent antinociceptive effects. After peripheral nerve injury the inhibitory control exerted by endogenous galanin is increased. During inflammation, galanin expression within the dorsal horn is increased. Endogenous galanin appears to play an enhanced antinociceptive role in chronic pain or neuropathic or inflammatory origin.

Galanin has been indicated in the etiology of depression. Galanin is colocalized within the serotoninergic and noradrenergic systems. An increase in the amount of galanin released from ascending noradrenergic neurons into the ventral tegmental area has been proposed to decrease dopamine release and thereby decrease motor activation and anhedonia, two major symptoms of depression. The receptors involved in these functions are not known.

Galanin has also been shown to control gastrointestinal and cardiovascular actions. For example, in the guinea pig ileum, galanin administration inhibits neurally induced smooth muscle contractility probably via its ability to reduce acetylcholine release. In addition, galanin inhibits somatostatin and gastrin release. Galanin also decreases blood flow following injection into the mesenteric arteriole, as well as sodium and chloride net absorption.

Galanin thus plays an important role in a large variety of physiological processes.

The effects of galanin are mediated via G-protein coupled receptors for which three types have been cloned, GALR1, GALR2 and GALR3 (*see, e.g.*, Howard *et al.*, *FEBS letter*, 405:285-290 (1997); Bloomquist *et al.*, *Biochem. Biophys. Res. Commun.* 243:474-479 (1998); WO 98/15570; WO 99/31130; WO 97/46681; WO

97/26853). For most of the biological processes regulated by galanin, the specific receptors involved in these functions are not known.

Identifying additional G protein-coupled receptors would allow insight into the role of the each receptor in the different biological processes in which GPCR-mediated signaling is involved. There is a strong need in the art for diagnostic and therapeutic tools for detection and treatment of the numerous diseases and disorders involving GPCR-mediated signaling. In addition, identifying additional receptors for galanin would allow insight into the role of the each receptor in the different biological processes in which galanin is involved. Moreover, there is a strong need in the art for diagnostic and therapeutic tools for detection and treatment of the numerous diseases and disorders involving galanin signaling. This invention addresses these and other needs.

SUMMARY OF THE INVENTION

The present invention provides polypeptides having at least 70%, 75%, 80%, 85%, 90%, 95% or more identity with the polypeptides encoded by the nucleic acid molecules having a nucleotide sequence selected from the group consisting of the sequences set forth in Table 1. In one embodiment, the polypeptides of the invention are encoded by a nucleic acid molecule having a nucleotide sequence selected from the group consisting of the sequences set forth in Table 1. In other embodiments, the polypeptides of the present invention comprise a region of 15 amino acids or more, optionally 30 amino acids or more, having at least 80%, preferably at least 85%, and most preferably 90% or more, identity with a region of 15 amino acids or more, optionally 30 amino acids or more, from a polypeptide encoded by a nucleic acid molecule having a nucleotide sequence selected from the group consisting of the sequences set forth in Table 1. In some embodiments, the nucleic acids molecules encoding the polypeptides of the invention are operably linked to a heterologous promoter. The present invention also provides expression vectors comprising the nucleic acid molecules encoding the polypeptides of the invention, as well as host cells comprising the expression vectors. In one embodiment, the host cell is a mammalian cell.

The present invention is also directed to nucleic acid probes that specifically hybridize with the nucleic acid molecules encoding the described polypeptides. The probes can be DNA or RNA. Antisense nucleic acid molecules that specifically hybridize to the nucleic acid sequences encoding the polypeptides of the invention are also provided.

In another aspect, antibodies that specifically bind to the polypeptides of the invention are also provided. The antibodies can be monoclonal or polyclonal.

The antibodies and nucleic acid probes described above can be used to detect the presence of the polypeptides of the invention or of the nucleic acid molecules encoding the described polypeptides. They can be used to diagnose a variety of diseases and disorders in which G protein-coupled receptors are involved, such as, *e.g.*, Alzheimer's disease, amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, COPD, Crohn's disease, depression, Duchenne muscular dystrophy, embryonal carcinoma, epilepsy, Ewing's sarcoma, glioblastoma multiform, Hodgkin's disease, lymphoma, lung adenocarcinoma, lung small cell carcinoma, macular degeneration, malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, multiple sclerosis, osteoarthritis, osteoporosis, osteosarcoma, ovarian carcinoma, pancreatic carcinoma, Parkinson's disease, prostate carcinoma, psoriasis, rhabdomyosarcoma, renal cell carcinoma, rheumatoid arthritis, schizophrenia, seminoma, squamous cell carcinoma, tuberculosis, thyroid carcinoma, tonsil, transitional carcinoma of the bladder, ulcerative colitis, *etc.*

The present invention is also directed to methods for identifying compounds that modulate the expression of one or more polypeptides of the invention, the methods comprising culturing a cell in the presence of a modulator to form a first cell culture, contacting RNA or cDNA from the first cell culture with at least one probe, each probe comprising a polynucleotide sequence encoding a polypeptide of the invention, and determining whether the amount of the probe(s) which hybridizes to the RNA or cDNA from the first cell culture is increased or decreased relative to the amount of the probe(s) which hybridizes to RNA or cDNA from a second cell culture grown in the absence of the modulator.

In addition, the present invention provides methods for identifying compounds that modulate the activity of one or more polypeptides of the invention, the methods comprising culturing cells expressing at least one polypeptide of interest in the presence of a compound, measuring the activity of the polypeptide(s) or second messenger activity and determining whether the activity is increased or decreased relative to the activity of the polypeptide(s) or second messenger activity from a second cell culture grown in the absence of the modulator.

The compounds identified using the methods of the present invention can be modulators, activators, repressors, agonists or antagonists and have therapeutic uses

for treating a variety of disorders and/or diseases in which G protein-coupled receptors have been implicated, such as, *e.g.*, Alzheimer's disease, amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, COPD, Crohn's disease, depression, Duchenne muscular dystrophy, embryonal carcinoma, epilepsy, Ewing's sarcoma, glioblastoma multiform, Hodgkin's disease, lymphoma, lung adenocarcinoma, lung small cell carcinoma, macular degeneration, malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, multiple sclerosis, osteoarthritis, osteoporosis, osteosarcoma, ovarian carcinoma, pancreatic carcinoma, Parkinson's disease, prostate carcinoma, psoriasis, rhabdomyosarcoma, renal cell carcinoma, rheumatoid arthritis, schizophrenia, seminoma, squamous cell carcinoma, tuberculosis, thyroid carcinoma, tonsil, transitional carcinoma of the bladder, ulcerative colitis, *etc.*

The present invention provides is directed to polypeptides having at least 80% identity, optionally at least 85% identity, with the polypeptide encoded by the nucleic acid molecule having the nucleotide sequence set forth in SEQ ID NO:1. In one embodiment, the polypeptide of the present invention is the polypeptide encoded by the sequence set forth in SEQ ID NO:1. In other embodiments, the polypeptides of the present invention comprise a region of 15 amino acids or more, optionally 30 amino acids or more, having at least 80%, preferably at least 85% and most preferably 90% or more identity with a region of 15 amino acids or more, optionally 30 amino acids or more, from the polypeptide encoded by the nucleic acid molecule having the nucleotide sequence set forth in SEQ ID NO:1. Vectors comprising the nucleic acids encoding the polypeptides of the invention, and host cells comprising the expression vectors are also provided. In some embodiments, the nucleic acid molecules encoding the polypeptides of the invention are operably linked to a heterologous promoter. In some embodiments, the host cell is a mammalian cell.

The present invention is also directed to nucleic acid probes that specifically hybridize with the nucleic acid molecules encoding the polypeptides of the invention. The probes can be DNA or RNA. Antisense nucleic acid molecules that specifically hybridize to the nucleic acid molecules encoding the polypeptides of the invention are also provided.

In another aspect, antibodies that specifically bind to the polypeptides of the invention are also provided. The antibodies can be monoclonal or polyclonal.

The nucleic acid probes and antibodies described above can be used to detect the presence of the nucleic acid molecules encoding the polypeptides of the invention. They can be used to diagnose a variety of diseases and disorders in which galanin is involved, such as, cognition and memory disorders, anorexia, hormonal release disorders, cardiovascular activity disorders, pain perception disorders, obesity, diabetes, Alzheimer's disease, *etc.*

The present invention is also directed to methods for identifying compounds that modulate the expression of the polypeptides of the invention, comprising culturing a cell in the presence of a modulator to form a first cell culture, contacting RNA or cDNA from the first cell culture with a probe which comprises a polynucleotide sequence encoding the polypeptide of the invention, and determining whether the amount of the probe which hybridizes to the RNA or cDNA from the first cell culture is increased or decreased relative to the amount of the probe which hybridizes to RNA or cDNA from a second cell culture grown in the absence of the modulator.

In addition, the present invention provides a method for identifying compounds that modulate the activity of the polypeptides of the invention, comprising culturing cells expressing the polypeptide of interest in the presence of a compound, measuring the activity of the polypeptide or second messenger activity and determining whether the activity is increased or decreased relative to the activity of the polypeptide or second messenger activity from a second cell culture grown in the absence of the modulator.

The compounds identified using the methods of the present invention can be modulators, activators, repressors, agonists or antagonists and have therapeutic uses for treating a variety of disorders and/or diseases in which galanin has been implicated. For example, compounds that decrease the expression (repressors) or activity (antagonists) of the polypeptides of the invention can be used, *e.g.*, to treat obesity, diabetes, hyperlipidemia, stroke, cognitive disorders, Alzheimer's disease, and/or endocrine disorders. Compounds that increase expression (activators) or activity (agonists) of the polypeptides of the invention can be used, for example, to treat anorexia and to decrease noniception.

DESCRIPTION OF THE SPECIFIC EMBODIMENTS

I. INTRODUCTION

The present invention is directed to novel G protein-coupled receptors (GPCRs) that are useful for treating and diagnosing a number of diseases and disorders, including, but not limited to, Alzheimer's disease, amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, COPD, Crohn's disease, depression, Duchenne muscular dystrophy, embryonal carcinoma, epilepsy, Ewing's sarcoma, glioblastoma multiform, Hodgkin's disease, lymphoma, lung adenocarcinoma, lung small cell carcinoma, macular degeneration, malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, multiple sclerosis, osteoarthritis, osteoporosis, osteosarcoma, ovarian carcinoma, pancreatic carcinoma, Parkinson's disease, prostate carcinoma, psoriasis, rhabdomyosarcoma, renal cell carcinoma, rheumatoid arthritis, schizophrenia, seminoma, squamous cell carcinoma, tuberculosis, thyroid carcinoma, tonsil, transitional carcinoma of the bladder, ulcerative colitis, *etc.* The present invention also provides methods for identifying modulators of G protein-coupled receptor-mediated signaling. Such modulators are useful for treating the above-listed and other diseases and disorders.

In some aspects, the present invention is directed to new galanin receptors that are useful for treating and diagnosing a number of diseases and disorders, including, but not limited to, Alzheimer's disease, learning and memory disorders, hormonal problems, fat metabolism disorders, feeding disorders, pain perception disorders, diabetes, depression, *etc.* The present invention also provides methods for identifying modulators of galanin signaling. Such modulators are useful for treating the above-listed and other diseases and disorders.

The invention provides novel G protein-coupled receptors, as well as vectors and cells to express these novel GPCRs, including, e.g., galanin receptors. Probes and antibodies that can be used to detect the GPCRs of the invention are also provided, as well as antisense polynucleotides. The probes and antibodies are useful for diagnostic purposes. In addition, the nucleic acids encoding the polypeptides of the invention, antisense polynucleotides and polypeptides of the invention are useful for gene therapy applications. The present invention also provides nucleic acid molecules encoding the polypeptides of the invention operably linked to a heterologous promoter that drives expression of the protein encoded by the nucleic acid sequence.

The invention further provides methods of screening for modulators, *e.g.*, activators, inhibitors, stimulators, enhancers, agonists, and antagonists, of these novel G protein-coupled receptors. Such modulators of the activity of the GPCRs are useful for

pharmacological and genetic modulation of the signaling pathways in which GPCRs are involved. These methods of screening can be used to identify high affinity agonists and antagonists of GPCRs' activity. These modulatory compounds can then be used in pharmaceutical industry to regulate G protein-coupled receptor-mediated signaling to treat a variety of diseases or disorders. Thus, the invention provides assays for GPCR-mediated signaling modulation, where the G protein-coupled receptors of the invention or other molecules located downstream of the G protein coupled receptor act as direct or indirect reporter molecules for the effect of modulators on GPCR-mediated signaling. G protein-coupled receptors can be used in assays, *e.g.*, to measure changes in ligand binding, transcription, signal transduction, receptor-ligand interactions, second messenger concentrations, *in vitro*, *in vivo*, and *ex vivo*.

In some embodiments, the present invention provides novel galanin receptors (GAL4), as well as vectors and cells to express the galanin receptors. Probes and antibodies that can be used to detect the galanin receptors of the invention are also provided, as well as antisense polynucleotides. The probes and antibodies are useful for diagnostic purposes. In addition, the nucleic acids encoding the polypeptides of the invention, antisense polynucleotides and polypeptides of the invention are useful for gene therapy applications.

In some aspects, the invention further provides methods of screening for modulators, *e.g.*, activators, inhibitors, stimulators, enhancers, agonists, and antagonists, of these novel galanin receptors. Such modulators of the activity of the galanin receptors are useful for pharmacological and genetic modulation of the galanin signaling pathways. These methods of screening can be used to identify high affinity agonists and antagonists of galanin receptors' activity. These modulatory compounds can then be used in pharmaceutical industry to regulate galanin signaling to treat a variety of diseases or disorders. Thus, the invention provides assays for galanin signaling modulation, where the galanin receptors of the invention or other molecules located downstream in the galanin signaling pathway act as direct or indirect reporter molecules for the effect of modulators on galanin signaling. Galanin receptors can be used in assays, *e.g.*, to measure changes in ligand binding, transcription, signal transduction, receptor-ligand interactions, second messenger concentrations, *in vitro*, *in vivo*, and *ex vivo*.

II. DEFINITIONS

"Amplification primers" are oligonucleotides comprising either natural or analog nucleotides that can serve as the basis for the amplification of a selected nucleic acid sequence. They include, for example, both polymerase chain reaction primers and ligase chain reaction oligonucleotides.

"Antibody" refers to a polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof which specifically bind and recognize an analyte (antigen). The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively.

An exemplary immunoglobulin (antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kD) and one "heavy" chain (about 50-70 kD). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain (V_L) and variable heavy chain (V_H) refer to these light and heavy chains respectively.

Antibodies exist, *e.g.*, as intact immunoglobulins or as a number of well characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce $F(ab)'_2$, a dimer of Fab which itself is a light chain joined to V_H-C_H1 by a disulfide bond. The $F(ab)'_2$ may be reduced under mild conditions to break the disulfide linkage in the hinge region, thereby converting the $F(ab)'_2$ dimer into an Fab' monomer. The Fab' monomer is essentially an Fab with part of the hinge region (*see*, Paul (Ed.) *Fundamental Immunology*, Third Edition, Raven Press, NY (1993)). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized *de novo* either chemically or by utilizing recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments either produced by the modification of whole antibodies or those synthesized *de novo* using recombinant DNA methodologies (*e.g.*, single chain Fv).

"Biological samples" refers to any tissue or liquid sample having genomic DNA or other nucleic acids (*e.g.*, mRNA) or proteins. It refers to samples of cells or tissue from a normal healthy individual as well as samples of cells or tissue from a subject

suspected of having, *e.g.*, Alzheimer's disease, rheumatoid arthritis, osteoarthritis, osteoporosis, amyotrophic lateral sclerosis, multiple sclerosis and atherosclerosis, asthma, depression, epilepsy, schizophrenia, Parkinson's disease, a sarcoma (*e.g.*, chondrosarcoma, Ewing's sarcoma, osteosarcoma, *etc.*), a carcinoma (*e.g.*, basal cell carcinoma, breast carcinoma, embryonal carcinoma, ovarian carcinoma, renal cell carcinoma, lung adenocarcinoma, lung small cell carcinoma, pancreatic carcinoma, prostate carcinoma, transitional carcinoma of the bladder, squamous cell carcinoma, thyroid carcinoma, *etc.*), psoriasis, cardiomyopathy, Crohn's disease, Duchenne muscular dystrophy, glioblastoma multiform, Hodgkin's disease, lymphoma, macular degeneration, malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, seminoma, tuberculosis, tonsil, ulcerative colitis, or any other disease or disorder in which G protein-coupled receptors are involved, as well as learning and/or memory disorders, diabetes, pain perception disorders, anorexia, obesity, hormonal release problems, or any other disease or disorder in which galanin is involved..

The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

The term "isolated," when applied to a nucleic acid or protein, denotes that the nucleic acid or protein is essentially free of other cellular components with which it is associated in the natural state. It is preferably in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein which is the predominant species present in a preparation is substantially purified. In particular, an isolated gene is separated from open reading frames which flank the gene and encode a protein other than the gene of interest. The term "purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or protein is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure.

The term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural nucleotides which have similar binding properties as the reference nucleic acid and are

metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.,* degenerate codon substitutions) and complementary sequences as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al., Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al., J. Biol. Chem.* 260:2605-2608 (1985); and Cassol *et al.* (1992); Rossolini *et al., Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, and mRNA encoded by a gene.

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers. As used herein, the terms encompass amino acid chains of any length, including full length proteins (*i.e.*, antigens), wherein the amino acid residues are linked by covalent peptide bonds.

The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, *i.e.*, an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, *e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. "Amino acid mimetics" refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, "conservatively modified variants" refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

The following eight groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Glycine (G);
2) Aspartic acid (D), Glutamic acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V);
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W);
7) Serine (S), Threonine (T); and

8) Cysteine (C), Methionine (M)

(*see, e.g.*, Creighton, *Proteins* (1984)).

Macromolecular structures such as polypeptide structures can be described in terms of various levels of organization. For a general discussion of this organization, *see, e.g.*, Alberts *et al.*, *Molecular Biology of the Cell* (3^{rd} ed., 1994) and Cantor and Schimmel, *Biophysical Chemistry Part I: The Conformation of Biological Macromolecules* (1980). "Primary structure" refers to the amino acid sequence of a particular peptide. "Secondary structure" refers to locally ordered, three dimensional structures within a polypeptide. These structures are commonly known as domains. Domains are portions of a polypeptide that form a compact unit of the polypeptide and are typically 50 to 350 amino acids long. Typical domains are made up of sections of lesser organization such as stretches of β-sheet and α-helices. "Tertiary structure" refers to the complete three dimensional structure of a polypeptide monomer. "Quaternary structure" refers to the three dimensional structure formed by the noncovalent association of independent tertiary units. Anisotropic terms are also known as energy terms.

"Percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (*i.e.*, 60% identity, optionally 65%, 70%, 75%, 80%, 85%, 90%, or 95% identity over a specified region), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be "substantially identical." This definition also refers to the complement of a test sequence. Optionally, the identity exists over a

region that is at least about 50 amino acids or nucleotides in length, or more preferably over a region that is 75-100 amino acids or nucleotides in length.

The term "similarity," or percent "similarity," in the context of two or more polypeptide sequences, refer to two or more sequences or subsequences that have a specified percentage of amino acid residues that are either the same or similar as defined in the 8 conservative amino acid substitutions defined above (*i.e.*, 60%, optionally 65%, 70%, 75%, 80%, 85%, 90%, or 95% similar over a specified region), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be "substantially similar." Optionally, this identity exists over a region that is at least about 50 amino acids in length, or more preferably over a region that is at least about 75-100 amino acids in length.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith and Waterman (1970) *Adv. Appl. Math.* 2:482c, by the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity method of Pearson and Lipman (1988) *Proc. Nat'l. Acad. Sci. USA* 85:2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575

Science Dr., Madison, WI), or by manual alignment and visual inspection (*see, e.g.*, Ausubel *et al.*, *Current Protocols in Molecular Biology* (1995 supplement)).

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It also plots a tree or dendogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng and Doolittle (1987) *J. Mol. Evol.* 35:351-360. The method used is similar to the method described by Higgins and Sharp (1989) *CABIOS* 5:151-153. The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters. Using PILEUP, a reference sequence is compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps. PILEUP can be obtained from the GCG sequence analysis software package, *e.g.*, version 7.0 (Devereaux *et al.* (1984) *Nuc. Acids Res.* 12:387-395).

Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.* (1977) *Nuc. Acids Res.* 25:3389-3402, and Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment

score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) or 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below. Yet another indication that two nucleic acid sequences are substantially identical is that the same primers can be used to amplify the sequence.

The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (*e.g.*, total cellular or library DNA or RNA).

The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10° C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (*e.g.*, 10 to 50 nucleotides) and at least about 60° C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, optionally 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5X SSC, and 1% SDS, incubating at 42°C, or 5X SSC, 1% SDS, incubating at 65°C, with wash in 0.2X SSC, and 0.1% SDS at 65°C. Such washes can be performed for 5, 15, 30, 60, 120, or more minutes.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions. Exemplary "moderately stringent hybridization conditions" include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 1X SSC at 45°C. Such

washes can be performed for 5, 15, 30, 60, 120, or more minutes. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

For PCR, a temperature of about 36°C is typical for low stringency amplification, although annealing temperatures may vary between about 32°C and 48°C depending on primer length. For high stringency PCR amplification, a temperature of about 62°C is typical, although high stringency annealing temperatures can range from about 50°C to about 65°C, depending on the primer length and specificity. Typical cycle conditions for both high and low stringency amplifications include a denaturation phase of 90°C - 95°C for 30 sec - 2 min., an annealing phase lasting 30 sec. - 2 min., and an extension phase of about 72°C for 1 - 2 min.

As used herein a "nucleic acid probe" is defined as a nucleic acid capable of binding to a target nucleic acid (*e.g.*, a nucleic acid encoding a galanin receptor) of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. As used herein, a probe may include natural (*i.e.*, A, G, C, or T) or modified bases (7-deazaguanosine, inosine, *etc.*). In addition, the bases in a probe may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization. Thus, for example, probes may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages. It will be understood by one of skill in the art that probes may bind target sequences lacking complete complementarity with the probe sequence depending upon the stringency of the hybridization conditions.

Nucleic acid probes can be DNA or RNA fragments. DNA fragments can be prepared, for example, by digesting plasmid DNA, or by use of PCR, or synthesized by either the phosphoramidite method described by Beaucage and Carruthers

A "labeled nucleic acid probe" is a nucleic acid probe that is bound, either covalently, through a linker, or through ionic, van der Waals or hydrogen bonds to a label such that the presence of the probe may be determined by detecting the presence of the label bound to the probe.

The phrase "a nucleic acid sequence encoding" refers to a nucleic acid which contains sequence information for a structural RNA such as rRNA, a tRNA, or the primary amino acid sequence of a specific protein or peptide, or a binding site for a trans-acting regulatory agent. This phrase specifically encompasses degenerate codons (*i.e.*, different codons which encode a single amino acid) of the native sequence or sequences which may be introduced to conform with codon preference in a specific host cell.

The term "recombinant" when used with reference, *e.g.*, to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (nonrecombinant) form of the cell or express native genes that are otherwise abnormally expressed, under-expressed or not expressed at all.

The term "heterologous" when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged to make a new functional nucleic acid, *e.g.*, a promoter from one source and a coding region from another source. Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (*e.g.*, a fusion protein).

A "promoter" is defined as an array of nucleic acid control sequences that direct transcription of a nucleic acid. As used herein, a promoter includes necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. A "constitutive" promoter is a promoter that is active under most environmental and developmental conditions. An "inducible" promoter is a promoter that is active under environmental or developmental regulation. The term "operably linked" refers to a functional linkage between a nucleic acid

expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

An "expression vector" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell. The expression vector can be part of a plasmid, virus, or nucleic acid fragment. Typically, the expression vector includes a nucleic acid to be transcribed operably linked to a promoter.

The phrase "specifically (or selectively) binds to an antibody" or "specifically (or selectively) immunoreactive with", when referring to a protein or peptide, refers to a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein and do not bind in a significant amount to other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, antibodies raised against a protein having an amino acid sequence encoded by any of the polynucleotides of the invention can be selected to obtain antibodies specifically immunoreactive with that protein and not with other proteins, except for polymorphic variants. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays, Western blots, or immunohistochemistry are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. *See*, Harlow and Lane *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, NY (1988) for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity. Typically, a specific or selective reaction will be at least twice the background signal or noise and more typically more than 10 to 100 times background.

"Inhibitors," "activators," and "modulators" of G protein-coupled receptors expression or of G protein-coupled receptors' activity are used to refer to inhibitory, activating, or modulating molecules, respectively, identified using *in vitro* and *in vivo* assays for G protein-coupled receptors expression or G protein-mediated signaling, *e.g.*, ligands, agonists, antagonists, and their homologs and mimetics. Inhibitors are compounds that, *e.g.*, inhibit expression of a G protein-coupled receptor or bind to, partially or totally block stimulation, decrease, prevent, delay activation,

inactivate, desensitize, or down-regulate the activity of a G protein-coupled receptor, *e.g.*, antagonists. Activators are compounds that, *e.g.*, induce or activate the expression of a G protein-coupled receptor or bind to, stimulate, increase, open, activate, facilitate, enhance activation, sensitize or up-regulate the activity of G protein-coupled receptors, *e.g.*, agonists. Modulators include compounds that, *e.g.*, alter the interaction of a receptor with extracellular proteins that bind activators or inhibitors, G proteins, and kinases. Modulators include genetically modified versions of G protein-coupled receptors, *e.g.*, with altered activity, as well as naturally occurring and synthetic ligands, antagonists, agonists, small chemical molecules and the like. Assays for inhibitors, activators and modulators include, *e.g.*, expressing a G protein-coupled receptor in cells or cell membranes, applying putative modulator compounds, in the presence or absence of a GPCR ligand (such as galanin, where appropriate) and then determining the functional effects on G protein-mediated signaling, as described above. Samples or assays comprising G protein-coupled receptors that are treated with a potential activator, inhibitor, or modulator are compared to control samples without the inhibitor, activator, or modulator to examine the extent of inhibition. Control samples (untreated with inhibitors) are assigned a relative G protein-coupled receptor activity value of 100%. Inhibition of a G protein-coupled receptor is achieved when the G protein-coupled receptor activity value relative to the control is about 80%, optionally 50% or 25-0%. Activation of a G protein-coupled receptor is achieved when the G protein-coupled receptor activity value relative to the control is 110%, optionally 150%, optionally 200-500%, or 1000-3000% higher.

III. GENERAL RECOMBINANT NUCLEIC ACIDS METHODS FOR USE WITH THE INVENTION

In numerous embodiments of the present invention, nucleic acids encoding the GPCRs of interest will be isolated and cloned using recombinant methods. Such embodiments are used, *e.g.*, to isolate GPCR-encoding polynucleotides for protein expression or during the generation of variants, derivatives, expression cassettes, or other sequences derived from GPCRs, to monitor GPCR gene expression, for the isolation or detection of GPCR sequences in different species, for diagnostic purposes in a patient, *e.g.*, to detect mutations in GPCRs, *etc.* In one embodiment, the nucleic acids of the invention are from any mammal, including, in particular, *e.g.*, a human, a rat, a mouse, *etc.*

In addition, recombinant expression of a GPCR of interest in eukaryotic cells, is useful for making cell membrane preparations that can be used for receptor binding assays. Receptor binding assays are used, in particular, for screening for modulators of the activity of GPCRs.

A. General Recombinant Nucleic Acids Methods

The numerous applications of the present invention involving the cloning, synthesis, maintenance, mutagenesis, and other manipulations of nucleic acid sequences can be performed using routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and Ausubel *et al.*, *Current Protocols in Molecular Biology* (1994).

Nucleotide sizes are given in either kilobases (kb) or base pairs (bp). These are estimates derived from agarose or acrylamide gel electrophoresis or, alternatively, from published DNA sequences.

Oligonucleotides that are not commercially available can be chemically synthesized according to the solid phase phosphoramidite triester method first described by Beaucage and Caruthers, *Tetrahedron Letts.* 22(20):1859-1862 (1981), using an automated synthesizer, as described in Needham Van Devanter *et al.*, *Nucleic Acids Res.* 12:6159-6168 (1984). Purification of oligonucleotides is, for example, by either native acrylamide gel electrophoresis or by anion-exchange HPLC as described in Pearson and Reanier, *J. Chrom.* 255:137-149 (1983).

The nucleic acids described here, or fragments thereof, can be used as hybridization probes for genomic or cDNA libraries to isolate the corresponding complete gene (including regulatory and promoter regions, exons and introns) or cDNAs, in particular cDNA clones corresponding to full-length transcripts. The probes may also be used to isolate other genes and cDNAs which have a high sequence similarity to the gene of interest or similar biological activity. Probes of this type preferably have at least 30 bases and may contain, for example, 50 or more bases.

The sequence of the cloned genes and synthetic oligonucleotides can be verified using the chemical degradation method of Maxam and Gilbert, *Methods in Enzymology* 65:499-560 (1980). The sequence can be confirmed after the assembly of the oligonucleotide fragments into the double-stranded DNA sequence using the method

of Maxam and Gilbert, *supra*, or the chain termination method for sequencing double-stranded templates of Wallace *et al., Gene* 16:21-26 (1981). Southern blot hybridization techniques can be carried out according to Southern *et al., J. Mol. Biol.* 98:503 (1975).

B. Cloning Methods for the Isolation of Nucleotide Sequences Encoding the Desired Proteins

In general, the nucleic acids encoding the subject proteins are cloned from DNA sequence libraries that are made to encode copy DNA (cDNA) or genomic DNA. The particular sequences can be located by hybridizing with an oligonucleotide probe, the sequence of which can be derived from the sequences provided herein (*e.g.*, the sequences set forth in Table 1), which provides a reference for PCR primers and defines suitable regions for isolating G protein-coupled receptors specific probes. Alternatively, where the sequence is cloned into an expression library, the expressed recombinant protein can be detected immunologically with antisera or purified antibodies made against the G protein-coupled receptor of interest.

Methods for making and screening genomic and cDNA libraries are well-known to those of skill in the art (*see, e.g.*, Gubler and Hoffman, *Gene* 25:263-269 (1983); Benton and Davis, *Science* 196:180-182 (1977); and Sambrook, *supra*).

Briefly, to make the cDNA library, one should choose a source that is rich in mRNA. The mRNA can then be made into cDNA, ligated into a recombinant vector, and transfected into a recombinant host for propagation, screening and cloning. For a genomic library, the DNA is extracted from a suitable tissue and either mechanically sheared or enzymatically digested to yield fragments of preferably about 5-100 kb. The fragments are then separated by gradient centrifugation from undesired sizes and are constructed in bacteriophage lambda vectors. These vectors and phage are packaged *in vitro*, and the recombinant phages are analyzed by plaque hybridization. Colony hybridization is carried out as generally described in Grunstein *et al., Proc. Natl. Acad. Sci. USA* 72:3961-3965 (1975).

An alternative method combines the use of synthetic oligonucleotide primers with polymerase extension on an mRNA or DNA template. Suitable primers can be designed from specific GPCRs, *e.g.*, the sequences described in Table 1. This polymerase chain reaction (PCR) method amplifies the nucleic acids encoding the protein of interest directly from mRNA, cDNA, genomic libraries or cDNA libraries. Restriction endonuclease sites can be incorporated into the primers. Polymerase chain reaction or

other *in vitro* amplification methods may also be useful, for example, to clone nucleic acids encoding specific proteins and express said proteins, to synthesize nucleic acids that will be used as probes for detecting the presence of mRNA encoding a G protein-coupled receptor of the invention in physiological samples, for nucleic acid sequencing, or for other purposes (*see*, U.S. Patent Nos. 4,683,195 and 4,683,202). Genes amplified by a PCR reaction can be purified, *e.g.*, from agarose gels, and cloned into an appropriate vector.

Appropriate primers and probes for identifying the genes encoding the G protein-coupled receptors of the invention from mammalian tissues can be derived from the sequences provided herein, in particular the sequences set forth in Table 1. For a general overview of PCR, *see*, Innis *et al.*, *PCR Protocols: A Guide to Methods and Applications*, *Academic Press*, San Diego (1990).

Synthetic oligonucleotides can be used to construct genes. This is done using a series of overlapping oligonucleotides, usually 40-120 bp in length, representing both the sense and anti-sense strands of the gene. These DNA fragments are then annealed, ligated and cloned.

A gene encoding a G protein-coupled receptor of the invention can be cloned using intermediate vectors before transformation into mammalian cells for expression. These intermediate vectors are typically prokaryote vectors or shuttle vectors. The proteins can be expressed in either prokaryotes, using standard methods well-known to those of skill in the art, or eukaryotes as described *infra*.

C. Expression in Eukaryotes

Standard eukaryotic transfection methods are used to produce eukaryotic cell lines, *e.g.*, yeast, insect, or mammalian cell lines, which express large quantities of the G protein-coupled receptors of the invention which are then purified using standard techniques (*see, e.g.*, Colley *et al.*, *J. Biol. Chem.* 264:17619-17622, (1989); and *Guide to Protein Purification*, in Vol. 182 of *Methods in Enzymology* (Deutscher ed., 1990)).

Transformations of eukaryotic cells are performed according to standard techniques as described by Morrison, *J. Bact.*, 132:349-351 (1977), or by Clark-Curtiss and Curtiss, *Methods in Enzymology*, 101:347-362 R. Wu *et al.* (Eds) Academic Press, NY (1983).

Any of the well-known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate

transfection, polybrene, protoplast fusion, electroporation, liposomes, microinjection, plasma vectors, viral vectors and any of the other well-known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (*see* Sambrook *et al.*, *supra*). It is only necessary that the particular genetic engineering procedure utilized be capable of successfully introducing at least one gene into the host cell which is capable of expressing the protein.

The particular eukaryotic expression vector used to transport the genetic information into the cell is not particularly critical. Any of the conventional vectors used for expression in eukaryotic cells may be used. Expression vectors containing regulatory elements from eukaryotic viruses are typically used. Suitable vectors for use in the present invention include, but are not limited to, SV40 vectors, vectors derived from bovine papilloma virus or from the Epstein Barr virus and baculovirus vectors, and any other vector allowing expression of proteins under the direction of the SV-40 later promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

The vectors usually include selectable markers which result in gene amplification, such as, *e.g.*, thymidine kinase, aminoglycoside phosphotransferase, hygromycin B phosphotransferase, xanthine-guanine phosphoribosyl transferase, CAD (carbamyl phosphate synthetase, aspartate transcarbamylase, and dihydroorotase), adenosine deaminase, dihydrofolate reductase, asparagine synthetase and ouabain selection. Alternatively, high yield expression systems not involving gene amplification are also suitable, such as, *e.g.*, using a baculovirus vector in insect cells, with a target protein encoding sequence under the direction of the polyhedrin promoter or other strong baculovirus promoters.

The expression vector of the present invention will typically contain both prokaryotic sequences that facilitate the cloning of the vector in bacteria as well as one or more eukaryotic transcription units that are expressed only in eukaryotic cells, such as mammalian cells. The vector may or may not comprise a eukaryotic replicon. If a eukaryotic replicon is present, then the vector is amplifiable in eukaryotic cells using the appropriate selectable marker. If the vector does not comprise a eukaryotic replicon, no episomal amplification is possible. Instead, the transfected DNA integrates into the genome of the transfected cell, where the promoter directs expression of the desired gene. The expression vector is typically constructed from elements derived from different, well

characterized viral or mammalian genes. For a general discussion of the expression of cloned genes in cultured mammalian cells, *see*, Sambrook *et al.*, *supra*, Ch. 16.

The prokaryotic elements that are typically included in the mammalian expression vector include a replicon that functions in *E. coli*, a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the plasmid to allow insertion of eukaryotic sequences. The particular antibiotic resistance gene chosen is not critical, any of the many resistance genes known in the art are suitable. The prokaryotic sequences are preferably chosen such that they do not interfere with the replication of the DNA in eukaryotic cells.

The expression vector contains a eukaryotic transcription unit or expression cassette that contains all the elements required for the expression of the DNA encoding the G protein-coupled receptors of interest in eukaryotic cells. A typical expression cassette contains a promoter operably linked to the DNA sequence encoding the G protein-coupled receptor and signals required for efficient polyadenylation of the transcript. The DNA sequence encoding the protein may typically be linked to a cleavable signal peptide sequence to promote secretion of the encoded protein by the transformed cell. Such signal peptides would include, among others, the signal peptides from tissue plasminogen activator, insulin, and neuron growth factor, and juvenile hormone esterase of *Heliothis virescens*. Additional elements of the cassette may include enhancers and, if genomic DNA is used as the structural gene, introns with functional splice donor and acceptor sites.

Eukaryotic promoters typically contain two types of recognition sequences, the TATA box and upstream promoter elements. The TATA box, located 25-30 base pairs upstream of the transcription initiation site, is thought to be involved in directing RNA polymerase to begin RNA synthesis. The other upstream promoter elements determine the rate at which transcription is initiated.

Enhancer elements can stimulate transcription up to 1,000 fold from linked homologous or heterologous promoters. Enhancers are active when placed downstream or upstream from the transcription initiation site. Many enhancer elements derived from viruses have a broad host range and are active in a variety of tissues (*see*, *Enhancers and Eukaryotic Expression*, Cold Spring Harbor Pres, Cold Spring Harbor, NY (1983)).

In the construction of the expression cassette, the promoter is preferably positioned at about the same distance from the heterologous transcription start site as it is

from the transcription start site in its natural setting. As is known in the art, some variation in this distance can, however, be accommodated without loss of promoter function.

In addition to a promoter sequence, the expression cassette should also contain a transcription termination region downstream of the structural gene to provide for efficient termination. The termination region may be obtained from the same gene as the promoter sequence or may be obtained from a different gene.

If the mRNA encoded by the structural gene is to be efficiently translated, polyadenylation sequences are also commonly added to the vector construct. Two distinct sequence elements are required for accurate and efficient polyadenylation: GU or U rich sequences located downstream from the polyadenylation site and a highly conserved sequence of six nucleotides, AAUAAA, located 11-30 nucleotides upstream. Termination and polyadenylation signals that are suitable for the present invention include those derived from SV40, or a partial genomic copy of a gene already resident on the expression vector.

In addition to the elements already described, the expression vector of the present invention may typically contain other specialized elements intended to increase the level of expression of cloned genes or to facilitate the identification of cells that carry the transfected DNA. For instance, a number of animal viruses contain DNA sequences that promote the extra chromosomal replication of the viral genome in permissive cell types. Plasmids bearing these viral replicons are replicated episomally as long as the appropriate factors are provided by genes either carried on the plasmid or with the genome of the host cell.

The cDNA encoding the protein of interest can be ligated to various expression vectors for use in transforming host cell cultures. The vectors typically contain gene sequences to initiate transcription and translation of the G protein-coupled receptor gene. These sequences need to be compatible with the selected host cell. In addition, the vectors preferably contain a marker to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or metallothionein. Additionally, a vector might contain a replicative origin.

Cells of mammalian origin are illustrative of cell cultures useful for the production of, for example, a G protein-coupled receptor of interest. Mammalian cell systems often will be in the form of monolayers of cells, although mammalian cell suspensions may also be used. Illustrative examples of mammalian cell lines include

VERO and HeLa cells, NIH 3T3, COS, Chinese hamster ovary (CHO), WI38, BHK, COS-7 or MDCK cell lines.

As indicated above, the vector, *e.g.*, a plasmid, which is used to transform the host cell, preferably contains DNA sequences to initiate transcription and sequences to control the translation of the gene sequence encoding the G protein-coupled receptor of interest. These sequences are referred to as expression control sequences. Illustrative expression control sequences are described, *e.g.*, in Berman *et al.*, *Science*, 222:524-527 (1983); Thomsen *et al.*, *Proc. Natl. Acad. Sci.* 81:659-663 (1984); and Brinster *et al.*, *Nature* 296:39-42 (1982). The cloning vector containing the expression control sequences is cleaved using restriction enzymes, adjusted in size as necessary or desirable and ligated with sequences encoding the G protein-coupled receptor by means well-known in the art.

When higher animal host cells are employed, polyadenylation or transcription terminator sequences from known mammalian genes need to be incorporated into the vector. An example of a terminator sequence is the polyadenylation sequence from the bovine growth hormone gene. Sequences for accurate splicing of the transcript may also be included. An example of a splicing sequence is the VP1 intron from SV40 (Sprague *et al.*, *J. Virol.* 45:773-781 (1983)).

Additionally, gene sequences to control replication in the host cell may be incorporated into the vector such as those found in bovine papilloma virus type-vectors (*see*, Saveria-Campo, *"Bovine Papilloma virus DNA a Eukaryotic Cloning Vector"* In: *DNA Cloning Vol.II: a Practical Approach* (Glover Ed.), IRL Press, Arlington, Virginia pp. 213-238 (1985)).

The transformed cells are cultured by means well-known in the art. For example, such means are published in *Biochemical Methods in Cell Culture and Virology*, Kuchler, Dowden, Hutchinson and Ross, Inc. (1977). The expressed protein is isolated from cells grown as suspensions or as monolayers. The latter are recovered by well-known mechanical, chemical or enzymatic means.

IV. PURIFICATION OF THE PROTEINS FOR USE WITH THE INVENTION

After expression, the proteins of the present invention can be purified to substantial purity by standard techniques, including selective precipitation with substances as ammonium sulfate, column chromatography, immunopurification methods, and other methods known to those of skill in the art (*see*, *e.g.*, Scopes *Protein*

Purification: Principles and Practice, Springer-Verlag, NY (1982); U.S. Patent No. 4,673,641; Ausubel *et al., supra*; and Sambrook *et al., supra*).

A number of conventional procedures can be employed when a recombinant protein is being purified. For example, proteins having established molecular adhesion properties can be reversibly fused to the subject protein. With the appropriate ligand, a G protein-coupled receptor of interest, for example, can be selectively adsorbed to a purification column and then freed from the column in a relatively pure form. The fused protein is then removed by enzymatic activity. Finally, the G protein-coupled receptors of the invention can be purified using immunoaffinity columns.

A. Purification of Proteins from Recombinant Bacteria

When recombinant proteins are expressed by the transformed bacteria in large amounts, typically after promoter induction, although expression can be constitutive, the proteins may form insoluble aggregates. There are several protocols that are suitable for purification of protein inclusion bodies. For example, purification of aggregate proteins (hereinafter referred to as inclusion bodies) typically involves the extraction, separation and/or purification of inclusion bodies by disruption of bacterial cells typically, but not limited to, by incubation in a buffer of about 100-150 μg/ml lysozyme and 0.1% Nonidet P40, a non-ionic detergent. The cell suspension can be ground using a Polytron grinder (Brinkman Instruments, Westbury, NY). Alternatively, the cells can be sonicated on ice. Alternate methods of lysing bacteria are described in Ausubel *et al.*, and Sambrook *et al.*, both *supra*, and will be apparent to those of skill in the art.

The cell suspension is generally centrifuged and the pellet containing the inclusion bodies resuspended in buffer which does not dissolve but washes the inclusion bodies, *e.g.,* 20 mM Tris-HCl (pH 7.2), 1 mM EDTA, 150 mM NaCl and 2% Triton-X 100, a non-ionic detergent. It may be necessary to repeat the wash step to remove as much cellular debris as possible. The remaining pellet of inclusion bodies may be resuspended in an appropriate buffer (*e.g.,* 20 mM sodium phosphate, pH 6.8, 150 mM NaCl). Other appropriate buffers will be apparent to those of skill in the art.

Following the washing step, the inclusion bodies are solubilized by the addition of a solvent that is both a strong hydrogen acceptor and a strong hydrogen donor (or a combination of solvents each having one of these properties). The proteins that

formed the inclusion bodies may then be renatured by dilution or dialysis with a compatible buffer. Suitable solvents include, but are not limited to, urea (from about 4 M to about 8 M), formamide (at least about 80%, volume/volume basis), and guanidine hydrochloride (from about 4 M to about 8 M). Some solvents which are capable of solubilizing aggregate-forming proteins, such as SDS (sodium dodecyl sulfate) and 70% formic acid, are inappropriate for use in this procedure due to the possibility of irreversible denaturation of the proteins, accompanied by a lack of immunogenicity and/or activity. Although guanidine hydrochloride and similar agents are denaturants, this denaturation is not irreversible and renaturation may occur upon removal (by dialysis, for example) or dilution of the denaturant, allowing re-formation of the immunologically and/or biologically active protein of interest. After solubilization, the protein can be separated from other bacterial proteins by standard separation techniques.

Alternatively, it is possible to purify proteins from bacteria periplasm. Where the protein is exported into the periplasm of the bacteria, the periplasmic fraction of the bacteria can be isolated by cold osmotic shock in addition to other methods known to those of skill in the art (*see,* Ausubel *et al., supra*). To isolate recombinant proteins from the periplasm, the bacterial cells are centrifuged to form a pellet. The pellet is resuspended in a buffer containing 20% sucrose. To lyse the cells, the bacteria are centrifuged and the pellet is resuspended in ice-cold 5 mM $MgSO_4$ and kept in an ice bath for approximately 10 minutes. The cell suspension is centrifuged and the supernatant decanted and saved. The recombinant proteins present in the supernatant can be separated from the host proteins by standard separation techniques well-known to those of skill in the art.

B. Standard Protein Separation Techniques For Purifying Proteins

1. Solubility Fractionation

Often as an initial step, and if the protein mixture is complex, an initial salt fractionation can separate many of the unwanted host cell proteins (or proteins derived from the cell culture media) from the recombinant protein of interest. The preferred salt is ammonium sulfate. Ammonium sulfate precipitates proteins by effectively reducing the amount of water in the protein mixture. Proteins then precipitate on the basis of their solubility. The more hydrophobic a protein is, the more likely it is to precipitate at lower ammonium sulfate concentrations. A typical protocol is to add saturated ammonium sulfate to a protein solution so that the resultant ammonium sulfate concentration is

between 20-30%. This will precipitate the most hydrophobic proteins. The precipitate is discarded (unless the protein of interest is hydrophobic) and ammonium sulfate is added to the supernatant to a concentration known to precipitate the protein of interest. The precipitate is then solubilized in buffer and the excess salt removed if necessary, through either dialysis or diafiltration. Other methods that rely on solubility of proteins, such as cold ethanol precipitation, are well-known to those of skill in the art and can be used to fractionate complex protein mixtures.

2. Size Differential Filtration

Based on a calculated molecular weight, a protein of greater and lesser size can be isolated using ultrafiltration through membranes of different pore sizes (for example, Amicon or Millipore membranes). As a first step, the protein mixture is ultrafiltered through a membrane with a pore size that has a lower molecular weight cut-off than the molecular weight of the protein of interest. The retentate of the ultrafiltration is then ultrafiltered against a membrane with a molecular cut off greater than the molecular weight of the protein of interest. The recombinant protein will pass through the membrane into the filtrate. The filtrate can then be chromatographed as described below.

3. Column Chromatography

The proteins of interest can also be separated from other proteins on the basis of their size, net surface charge, hydrophobicity and affinity for ligands. In addition, antibodies raised against proteins can be conjugated to column matrices and the proteins immunopurified. All of these methods are well-known in the art.

It will be apparent to one of skill that chromatographic techniques can be performed at any scale and using equipment from many different manufacturers (*e.g.*, Pharmacia Biotech).

V. DETECTION OF GENE EXPRESSION OF THE GPCRs

The polypeptides of the present invention and the polynucleotides encoding them can be employed as research reagents and materials for discovery of treatments and diagnostics to human disease. It will be readily apparent to those of skill in the art that although the following discussion is directed to methods for detecting nucleic acids encoding a G protein-coupled receptor, similar methods can be used to detect nucleic acids associated with, *e.g.*, Alzheimer's disease, depression, specific carcinomas and sarcomas, or any disease or disorder in which GPCR-mediated signaling

is involved. In aspects involving, e.g., a galanin receptor, similar methods can be used to detect nucleic acids associated with, *e.g.*, Alzheimer's disease, learning and memory disorders, reproduction and sex behavior disorders, feeding disorders, fat metabolism and body adiposity, regulation of neurotransmitter release, pain perception, depression, regulation of hormone release, cardiovascular actions regulation, or any disease or disorder in which galanin signaling is involved.

As should be apparent to those of skill in the art, the invention is based, at least in part, in the identification of novel G protein-coupled receptors, including a novel galanin receptor (GAL4). Accordingly, the present invention also includes methods for detecting the presence, alteration or absence of nucleic acids (*e.g.*, DNA or RNA) encoding such G protein-coupled receptors in a physiological specimen in order to determine the presence of, *e.g.*, Alzheimer's disease, amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, COPD, Crohn's disease, depression, Duchenne muscular dystrophy, embryonal carcinoma, epilepsy, Ewing's sarcoma, glioblastoma multiform, Hodgkin's disease, lymphoma, lung adenocarcinoma, lung small cell carcinoma, macular degeneration, malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, multiple sclerosis, osteoarthritis, osteoporosis, osteosarcoma, ovarian carcinoma, pancreatic carcinoma, Parkinson's disease, prostate carcinoma, psoriasis, rhabdomyosarcoma, renal cell carcinoma, rheumatoid arthritis, schizophrenia, seminoma, squamous cell carcinoma, tuberculosis, thyroid carcinoma, tonsil, transitional carcinoma of the bladder, ulcerative colitis, *etc.*, associated with mutations created in the sequences encoding the GPCRs that modify the expression and/or activity of the receptors, including those disorders aassociated with mutations created in the sequences encoding the galanin receptor that modify the activity of the receptor, including cognitive deficit, Alzheimer's disease, reproductive disorder, fat metabolism disorder, inhibition of neurotransmitter release, pain perception disorder, depression, hormone release disorder, decrease in blood flow, *etc.* Any tissue having cells bearing the genome of an individual, or RNA encoding the GPCRs can be used as well as biopsies of suspect tissue. It is also possible and preferred in some circumstances to conduct assays on cells that are isolated under microscopic visualization. A particularly useful method is the microdissection technique described in WO 95/23960. The cells isolated by microscopic visualization can be used in any of the assays described herein including both genomic and immunological based assays.

This invention provides methods of genotyping family members in which relatives are diagnosed with, *e.g.*, Alzheimer's disease, amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, COPD, Crohn's disease, depression, Duchenne muscular dystrophy, embryonal carcinoma, epilepsy, Ewing's sarcoma, glioblastoma multiform, Hodgkin's disease, lymphoma, lung adenocarcinoma, lung small cell carcinoma, macular degeneration, malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, multiple sclerosis, osteoarthritis, osteoporosis, osteosarcoma, ovarian carcinoma, pancreatic carcinoma, Parkinson's disease, prostate carcinoma, psoriasis, rhabdomyosarcoma, renal cell carcinoma, rheumatoid arthritis, schizophrenia, seminoma, squamous cell carcinoma, tuberculosis, thyroid carcinoma, tonsil, transitional carcinoma of the bladder, ulcerative colitis, Alzheimer's disease, depression, fat metabolism disorders, anorexia, stroke, diabetes, *etc.* Conventional methods of genotyping are known to those of skill in the art.

The probes are capable of binding to a target nucleic acid (*e.g.*, a nucleic acid encoding a G protein-coupled receptor of interest). By assaying for the presence or absence of the probe, one can detect the presence or absence of the target nucleic acid in a sample. Preferably, non-hybridizing probe and target nucleic acids are removed (*e.g.*, by washing) prior to detecting the presence of the probe.

A variety of methods of specific DNA and RNA measurement using nucleic acid hybridization techniques are known to those of skill in the art (*see*, Sambrook, *supra*). Some methods involve an electrophoretic separation (*e.g.*, Southern blot for detecting DNA, and Northern blot for detecting RNA), but measurement of DNA and RNA can also be carried out in the absence of electrophoretic separation (*e.g.*, by dot blot). Southern blot of genomic DNA (*e.g.*, from a human) can be used for screening for restriction fragment length polymorphism (RFLP) to detect the presence of a genetic disorder affecting a G protein-coupled receptor of the invention.

The selection of a nucleic acid hybridization format is not critical. A variety of nucleic acid hybridization formats are known to those skilled in the art. For example, common formats include sandwich assays and competition or displacement assays. Hybridization techniques are generally described in Hames and Higgins, *Nucleic Acid Hybridization, A Practical Approach*, IRL Press (1985); Gall and Pardue, *Proc. Natl. Acad. Sci. U.S.A.*, 63:378-383 (1969); and John *et al.*, *Nature*, 223:582-587 (1969).

Detection of a hybridization complex may require the binding of a signal generating complex to a duplex of target and probe polynucleotides or nucleic acids. Typically, such binding occurs through ligand and anti-ligand interactions as between a ligand-conjugated probe and an anti-ligand conjugated with a signal. The binding of the signal generation complex is also readily amenable to accelerations by exposure to ultrasonic energy.

The label may also allow indirect detection of the hybridization complex. For example, where the label is a hapten or antigen, the sample can be detected by using antibodies. In these systems, a signal is generated by attaching fluorescent or enzyme molecules to the antibodies or in some cases, by attachment to a radioactive label (*see, e.g.*, Tijssen, "*Practice and Theory of Enzyme Immunoassays*," *Laboratory Techniques in Biochemistry and Molecular Biology*, pp. 9-20, Burdon and van Knippenberg Eds., Elsevier (1985)).

The probes are typically labeled either directly, as with isotopes, chromophores, lumiphores, chromogens, or indirectly, such as with biotin, to which a streptavidin complex may later bind. Thus, the detectable labels used in the assays of the present invention can be primary labels (where the label comprises an element that is detected directly or that produces a directly detectable element) or secondary labels (where the detected label binds to a primary label, *e.g.*, as is common in immunological labeling). Typically, labeled signal nucleic acids are used to detect hybridization. Complementary nucleic acids or signal nucleic acids may be labeled by any one of several methods typically used to detect the presence of hybridized polynucleotides. The most common method of detection is the use of autoradiography with ^{3}H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P-labeled probes or the like.

Other labels include, *e.g.*, ligands which bind to labeled antibodies, fluorophores, chemiluminescent agents, enzymes, and antibodies which can serve as specific binding pair members for a labeled ligand. An introduction to labels, labeling procedures and detection of labels is found in Polak and Van Noorden, *Introduction to Immunocytochemistry*, 2nd ed., Springer Verlag, NY (1997); and in Haugland, *Handbook of Fluorescent Probes and Research Chemicals*, a combined handbook and catalogue Published by Molecular Probes, Inc. (1996).

In general, a detector which monitors a particular probe or probe combination is used to detect the detection reagent label. Typical detectors include spectrophotometers, phototubes and photodiodes, microscopes, scintillation counters,

cameras, film and the like, as well as combinations thereof. Examples of suitable detectors are widely available from a variety of commercial sources known to persons of skill in the art. Commonly, an optical image of a substrate comprising bound labeling moieties is digitized for subsequent computer analysis.

Most typically, the amount of, for example, a G protein-coupled receptor RNA is measured by quantitating the amount of label fixed to the solid support by binding of the detection reagent. Typically, the presence of a modulator during incubation will increase or decrease the amount of label fixed to the solid support relative to a control incubation which does not comprise the modulator, or as compared to a baseline established for a particular reaction type. Means of detecting and quantitating labels are well-known to those of skill in the art.

In preferred embodiments, the target nucleic acid or the probe is immobilized on a solid support. Solid supports suitable for use in the assays of the invention are known to those of skill in the art. As used herein, a solid support is a matrix of material in a substantially fixed arrangement.

A variety of automated solid-phase assay techniques are also appropriate. For instance, very large scale immobilized polymer arrays (VLSIPS™), available from Affymetrix, Inc. in Santa Clara, CA, can be used to detect changes in expression levels of a plurality of genes involved in the same regulatory pathways simultaneously. *See*, Tijssen, *supra.*, Fodor *et al.*, *Science*, 251:767-777 (1991); Sheldon *et al.*, *Clinical Chemistry* 39(4):718-719 (1993); and Kozal *et al.*, *Nature Medicine* 2(7):753-759 (1996). Thus, in one embodiment, the invention provides methods of detecting expression levels of the G protein-coupled receptors of the invention in combination with other G protein-coupled receptors and other nucleic acids known to be involved in regulating, *e.g.*, Alzheimer's disease, depression, feeding behavior, diabetes, obesity, stroke, cognition and memory, hormone release, amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, COPD, Crohn's disease, depression, Duchenne muscular dystrophy, embryonal carcinoma, epilepsy, Ewing's sarcoma, glioblastoma multiform, Hodgkin's disease, lymphoma, lung adenocarcinoma, lung small cell carcinoma, macular degeneration, malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, multiple sclerosis, osteoarthritis, osteoporosis, osteosarcoma, ovarian carcinoma, pancreatic carcinoma, Parkinson's disease, prostate carcinoma, psoriasis, rhabdomyosarcoma, renal cell carcinoma, rheumatoid arthritis, schizophrenia, seminoma, squamous cell carcinoma, tuberculosis,

thyroid carcinoma, tonsil, transitional carcinoma of the bladder, ulcerative colitis, *etc.*, in which nucleic acids (*e.g.,* RNA from a cell culture) are hybridized to an array of nucleic acids that are known to be associated with the above-listed diseases and disorders. Thus, in one embodiment, the invention provides methods for detecting the expression levels of nucleic acids encoding the G protein-coupled receptors of the invention, in which nucleic acids (*e.g.,* RNA from a cell culture) are hybridized to an array of nucleic acids that are known to be associated with the above-listed diseases and disorders in which GPCRs have been implicated. In a second embodiment, the invention provides methods for detecting the expression levels of nucleic acids encoding the galanin receptors of the invention, in which nucleic acids (*e.g.,* RNA from a cell culture) are hybridized to an array of nucleic acids that are known to be associated with Alzheimer's disease, depression, fat metabolism disorders, feeding disorders, hormonal disorders, *etc.* For example, in the assay described *supra*, oligonucleotides which hybridize to a plurality of nucleic acids encoding either G protein-coupled receptors or other molecules known to be involved in the above-mentioned diseases and disorders are optionally synthesized on a DNA chip (such chips are available from Affymetrix) and the RNA from a biological sample, such as a cell culture, is hybridized to the chip for simultaneous analysis of multiple nucleic acids. The nucleic acids encoding the G protein-coupled receptors that are present in the sample which is assayed are detected at specific positions on the chip.

Detection can be accomplished, for example, by using a labeled detection moiety that binds specifically to duplex nucleic acids (*e.g.,* an antibody that is specific for RNA-DNA duplexes). One preferred example uses an antibody that recognizes DNA-RNA heteroduplexes in which the antibody is linked to an enzyme (typically by recombinant or covalent chemical bonding). The antibody is detected when the enzyme reacts with its substrate, producing a detectable product. Coutlee *et al.*, *Analytical Biochemistry* 181:153-162 (1989); Bogulavski *et al.*, *J. Immunol. Methods* 89:123-130 (1986); Prooijen-Knegt, *Exp. Cell Res.* 141:397-407 (1982); Rudkin, *Nature* 265:472-473 (1976); Stollar, *PNAS* 65:993-1000 (1970); Ballard, *Mol. Immunol.* 19:793-799 (1982); Pisetsky and Caster, *Mol. Immunol.* 19:645-650 (1982); Viscidi *et al.*, *J. Clin. Microbial.* 41:199-209 (1988); and Kiney *et al.*, *J. Clin. Microbiol.* 27:6-12 (1989) describe antibodies to RNA duplexes, including homo and heteroduplexes. Kits comprising antibodies specific for DNA:RNA hybrids are available, *e.g.*, from Digene Diagnostics, Inc. (Beltsville, MD).

In addition to available antibodies, one of skill in the art can easily make antibodies specific for nucleic acid duplexes using existing techniques, or modify those antibodies which are commercially or publicly available. In addition to the art referenced above, general methods for producing polyclonal and monoclonal antibodies are known to those of skill in the art (*see, e.g.*, Paul (ed), *Fundamental Immunology, Third Edition* Raven Press, Ltd., NY (1993); Coligan, *Current Protocols in Immunology* Wiley/Greene, NY (1991); Harlow and Lane, *Antibodies: A Laboratory Manual* Cold Spring Harbor Press, NY (1989); Stites *et al.* (eds.), *Basic and Clinical Immunology* (4th ed.) Lange Medical Publications, Los Altos, CA, and references cited therein; Goding, *Monoclonal Antibodies: Principles and Practice* (2d ed.) Academic Press, New York, NY, (1986); and Kohler and Milstein, *Nature* 256:495-497 (1975)). Other suitable techniques for antibody preparation include selection of libraries of recombinant antibodies in phage or similar vectors (*see*, Huse *et al.*, *Science* 246:1275-1281 (1989); and Ward *et al.*, *Nature* 341:544-546 (1989)). Specific monoclonal and polyclonal antibodies and antisera will usually bind with a K_D of at least about 0.1 μM, preferably at least about 0.01 μM or better, and most typically and preferably, 0.001 μM or better.

The nucleic acids used in this invention can be either positive or negative probes. Positive probes bind to their targets and the presence of duplex formation is evidence of the presence of the target. Negative probes fail to bind to the suspect target and the absence of duplex formation is evidence of the presence of the target. For example, the use of a wild type specific nucleic acid probe or PCR primers may serve as a negative probe in an assay sample where only the nucleotide sequence of interest is present.

The sensitivity of the hybridization assays may be enhanced through use of a nucleic acid amplification system which multiplies the target nucleic acid being detected. Examples of such systems include the polymerase chain reaction (PCR) system and the ligase chain reaction (LCR) system. Other methods recently described in the art are the nucleic acid sequence based amplification (NASBA9, Cangene, Mississauga, Ontario) and Q Beta Replicase systems. These systems can be used to directly identify mutants where the PCR or LCR primers are designed to be extended or ligated only when a selected sequence is present. Alternatively, the selected sequences can be generally amplified using, for example, nonspecific PCR primers and the amplified target region later probed for a specific sequence indicative of a mutation.

A preferred embodiment is the use of allelic specific amplifications. In the case of PCR, the amplification primers are designed to bind to a portion of, for example, a gene encoding a G protein-coupled receptor protein, but the terminal base at the 3' end is used to discriminate between the mutant and wild-type forms of the G protein-coupled receptor gene. If the terminal base matches the point mutation or the wild-type, polymerase dependent three prime extension can proceed and an amplification product is detected. This method for detecting point mutations or polymorphisms is described in detail by Sommer *et al.*, in *Mayo Clin. Proc.* 64:1361-1372 (1989). By using appropriate controls, one can develop a kit having both positive and negative amplification products. The products can be detected using specific probes or by simply detecting their presence or absence. A variation of the PCR method uses LCR where the point of discrimination, *i.e.*, either the point mutation or the wild-type bases fall between the LCR oligonucleotides. The ligation of the oligonucleotides becomes the means for discriminating between the mutant and wild-type forms of the gene encoding the G protein-coupled receptor.

An alternative means for determining the level of expression of the nucleic acids of the present invention is *in situ* hybridization. *In situ* hybridization assays are well-known and are generally described in Angerer *et al.*, *Methods Enzymol.* 152:649-660 (1987). In an *in situ* hybridization assay, cells, preferentially human cells from the cerebellum or the hippocampus, are fixed to a solid support, typically a glass slide. If DNA is to be probed, the cells are denatured with heat or alkali. The cells are then contacted with a hybridization solution at a moderate temperature to permit annealing of specific probes that are labeled. The probes are preferably labeled with radioisotopes or fluorescent reporters.

VI. IMMUNOLOGICAL DETECTION OF THE GPCRs

In numerous embodiments of the present invention, antibodies that specifically bind to the G protein-coupled receptors of the invention will be used. Such antibodies have numerous applications, including for the modulation of the activity of the G protein-coupled receptors and for immunoassays to detect the G protein-coupled receptors of the invention, as well as variants, derivatives, fragments, *etc.* thereof. Immunoassays can be used to qualitatively or quantitatively analyze the proteins of interest. A general overview of the applicable technology can be found in Harlow and Lane, *Antibodies: A Laboratory Manual,* Cold Spring Harbor Pubs., NY (1988).

Immunoassays for detecting target G protein-coupled receptor proteins are useful for diagnosing any disease or disorder in which GPCR-mediated signaling has been involved such as, *e.g.*, Alzheimer's disease, depression, specific sarcomas and carcinomas, Parkinson's disease, psoriasis, rheumatoid arthritis, schizophrenia, tuberculosis, learning and memory disorders, diabetes, reproduction and sex behavior disorders, anorexia, fat metabolism and body adiposity disorders, regulation of neurotransmitter release, pain perception, depression, regulation of hormone release, cardiovascular actions regulation, *etc.* In some embodiments, the antibodies of the present invention specifically bind to the G protein-coupled receptors of the invention and do not bind to other G protein-coupled receptors or to G protein-coupled receptors from a different species, such as mouse, rat, *etc.* (identified GPCRs are listed in public databases, such as SwissProt, *see* http://www.expasy.ch/sprot/sprot-top.html, or GenBank, *see* http://www.ncbi.nlm.nih.gov/; *see* also *G protein coupled receptor Database*, http://www.gcrdb.uthscsa.edu). In some embodiments, the antibodies of the present invention specifically bind to the galanin receptors of the invention and do not bind to other galanin receptors, such as GALR1, GALR2 and GALR3 (*see, e.g.*, SwissProt accession numbers P47211, O43603, and O60755 for the sequences of the human GALR1, GALR2 and GALR3, respectively) or to galanin receptors from a different species (*see, e.g.*, SwissProt accession numbers P56479, O88854, O88853, for the sequences of the mouse GALR1, GALR2, and GALR3, respectively, and accession numbers Q62805, O08726, and O88626, for the sequences of the rat GALR1, GALR2, and GALR3, respectively).

A. Antibodies to Target Proteins

Methods for producing polyclonal and monoclonal antibodies that react specifically with a protein of interest are known to those of skill in the art (*see, e.g.*, Coligan, *supra*; and Harlow and Lane, *supra*; Stites *et al.*, *supra* and references cited therein; Goding, *supra*; and Kohler and Milstein, *Nature* 256:495-497 (1975)). Such techniques include antibody preparation by selection of antibodies from libraries of recombinant antibodies in phage or similar vectors (*see*, Huse *et al.*, *supra*; and Ward *et al.*, *supra*). For example, in order to produce antisera for use in an immunoassay, the protein of interest or an antigenic fragment thereof, is isolated as described herein. For example, a recombinant protein is produced in a transformed cell line. An inbred strain of mice or rabbits is immunized with the protein using a standard adjuvant, such as

Freund's adjuvant, and a standard immunization protocol. Alternatively, a synthetic peptide derived from the sequences disclosed herein and conjugated to a carrier protein can be used as an immunogen.

Polyclonal sera are collected and titered against the immunogen protein in an immunoassay, for example, a solid phase immunoassay with the immunogen immobilized on a solid support. Polyclonal antisera with a titer of 10^4 or greater are selected and tested for their cross-reactivity against non-G protein-coupled receptor proteins or even other homologous proteins from other organisms, using a competitive binding immunoassay. Specific monoclonal and polyclonal antibodies and antisera will usually bind with a K_D of at least about 0.1 mM, more usually at least about 1 μM, preferably at least about 0.1 μM or better, and most preferably, 0.01 μM or better.

A number of proteins of the invention comprising immunogens may be used to produce antibodies specifically or selectively reactive with the proteins of interest. Recombinant protein is the preferred immunogen for the production of monoclonal or polyclonal antibodies. Naturally occurring protein may also be used either in pure or impure form. Synthetic peptides made using the protein sequences described herein may also be used as an immunogen for the production of antibodies to the protein. Recombinant protein can be expressed in eukaryotic or prokaryotic cells and purified as generally described *supra*. The product is then injected into an animal capable of producing antibodies. Either monoclonal or polyclonal antibodies may be generated for subsequent use in immunoassays to measure the protein.

Methods of production of polyclonal antibodies are known to those of skill in the art. In brief, an immunogen, preferably a purified protein, is mixed with an adjuvant and animals are immunized. The animal's immune response to the immunogen preparation is monitored by taking test bleeds and determining the titer of reactivity to the G protein-coupled receptor of interest. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the protein can be done if desired (*see*, Harlow and Lane, *supra*).

Monoclonal antibodies may be obtained using various techniques familiar to those of skill in the art. Typically, spleen cells from an animal immunized with a desired antigen are immortalized, commonly by fusion with a myeloma cell (*See*, Kohler and Milstein, *Eur. J. Immunol.* 6:511-519 (1976)). Alternative methods of immortalization include, *e.g.*, transformation with Epstein Barr Virus, oncogenes, or

retroviruses, or other methods well-known in the art. Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Alternatively, one may isolate DNA sequences which encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according to the general protocol outlined by Huse *et al.*, *supra*.

Once target protein specific antibodies are available, the protein can be measured by a variety of immunoassay methods with qualitative and quantitative results available to the clinician. For a review of immunological and immunoassay procedures in general, *see*, Stites, *supra*. Moreover, the immunoassays of the present invention can be performed in any of several configurations, which are reviewed extensively in Maggio, *Enzyme Immunoassay*, CRC Press, Boca Raton, Florida (1980); Tijssen, *supra*; and Harlow and Lane, *supra*.

Immunoassays to measure target proteins in a human sample may use a polyclonal antiserum which was raised to the protein partially encoded by a sequence described herein (*e.g.*, a sequence selected from the sequences set forth in Table 1) or a fragment thereof. This antiserum is selected to have low cross-reactivity against non-G protein-coupled receptor proteins and any such cross-reactivity is removed by immunoabsorption prior to use in the immunoassay.

Polyclonal antibodies that specifically bind to a G protein-coupled receptor of interest from a particular species can be made by subtracting out cross-reactive antibodies using G protein-coupled receptor homologs. In an analogous fashion, antibodies specific to a particular G protein-coupled receptor (*e.g.*, a G protein-coupled receptor encoded by a sequence set forth in Table 1) can be obtained in an organism with multiple G protein-coupled receptors genes by subtracting out cross-reactive antibodies using other G protein-coupled receptors.

Polyclonal antibodies that specifically bind to a galanin receptor of interest from a particular species can be made by subtracting out cross-reactive antibodies using galanin receptor homologs. In an analogous fashion, antibodies specific to a particular galanin receptor (*e.g.*, the galanin receptors of the invention) can be obtained in an organism with multiple galanin receptor genes by subtracting out cross-reactive antibodies using other galanin receptors, such as GALR1, GALR2 and GALR3.

B. Immunological Binding Assays

In a preferred embodiment, a protein of interest is detected and/or quantified using any of a number of well-known immunological binding assays (*see, e.g.*, U.S. Patent Nos. 4,366,241; 4,376,110; 4,517,288; and 4,837,168). For a review of the general immunoassays, *see* also Asai, *Methods in Cell Biology Volume 37: Antibodies in Cell Biology*, Academic Press, Inc. NY (1993); Stites, *supra*. Immunological binding assays (or immunoassays) typically utilize a "capture agent" to specifically bind to and often immobilize the analyte (in this case a G protein-coupled receptor of the invention or antigenic subsequences thereof). The capture agent is a moiety that specifically binds to the analyte. In a preferred embodiment, the capture agent is an antibody that specifically binds, for example, a GPCR of the invention. The antibody (*e.g.*, anti-GPCR antibody) may be produced by any of a number of means well-known to those of skill in the art and as described above.

Immunoassays also often utilize a labeling agent to specifically bind to and label the binding complex formed by the capture agent and the analyte. The labeling agent may itself be one of the moieties comprising the antibody/analyte complex. Thus, the labeling agent may be a labeled GPCR polypeptide or a labeled anti-GPCR antibody. Alternatively, the labeling agent may be a third moiety, such as another antibody, that specifically binds to the antibody/protein complex.

In a preferred embodiment, the labeling agent is a second antibody bearing a label. Alternatively, the second antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second antibody can be modified with a detectable moiety, such as biotin, to which a third labeled molecule can specifically bind, such as enzyme-labeled streptavidin.

Other proteins capable of specifically binding immunoglobulin constant regions, such as protein A or protein G, can also be used as the label agents. These proteins are normal constituents of the cell walls of streptococcal bacteria. They exhibit a strong non-immunogenic reactivity with immunoglobulin constant regions from a variety of species (*see*, generally, Kronval *et al. J. Immunol.* 111:1401-1406 (1973); and Akerstrom *et al., J. Immunol.* 135:2589-2542 (1985)).

Throughout the assays, incubation and/or washing steps may be required after each combination of reagents. Incubation steps can vary from about 5 seconds to several hours, preferably from about 5 minutes to about 24 hours. The incubation time

will depend upon the assay format, analyte, volume of solution, concentrations, and the like. Usually, the assays will be carried out at ambient temperature, although they can be conducted over a range of temperatures, such as 10°C to 40°C.

1. Non-competitive Assay Formats

Immunoassays for detecting proteins of interest from tissue samples may be either competitive or noncompetitive. Noncompetitive immunoassays are assays in which the amount of captured analyte (in this case the protein) is directly measured. In one preferred "sandwich" assay, for example, the capture agent (*e.g.,* anti-GPCR antibodies) can be bound directly to a solid substrate where it is immobilized. These immobilized antibodies then capture the G protein-coupled receptor present in the test sample. The G protein-coupled receptor thus immobilized is then bound by a labeling agent, such as a second anti-GPCR antibody bearing a label. Alternatively, the second antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second can be modified with a detectable moiety, such as biotin, to which a third labeled molecule can specifically bind, such as enzyme-labeled streptavidin.

2. Competitive Assay Formats

In competitive assays, the amount of target protein (analyte) present in the sample is measured indirectly by measuring the amount of an added (exogenous) analyte (*i.e.,* a GPCR of interest) displaced (or competed away) from a capture agent (*i.e.*, anti-GPCR antibody) by the analyte present in the sample. In one competitive assay, a known amount of, in this case, the protein of interest is added to the sample and the sample is then contacted with a capture agent, in this case an antibody that specifically binds to the GPCR of interest. The amount of GPCR bound to the antibody is inversely proportional to the concentration of GPCR present in the sample. In a particularly preferred embodiment, the antibody is immobilized on a solid substrate. The amount of the GPCR bound to the antibody may be determined either by measuring the amount of subject protein present in a GPCR protein/antibody complex or, alternatively, by measuring the amount of remaining uncomplexed protein. The amount of GPCR protein may be detected by providing a labeled GPCR protein molecule.

A hapten inhibition assay is another preferred competitive assay. In this assay, a known analyte, in this case the target protein, is immobilized on a solid substrate. A known amount of anti-GPCR antibody is added to the sample, and the sample is then contacted with the immobilized target. In this case, the amount of anti-GPCR antibody

bound to the immobilized GPCR is inversely proportional to the amount of GPCR protein present in the sample. Again, the amount of immobilized antibody may be detected by detecting either the immobilized fraction of antibody or the fraction of the antibody that remains in solution. Detection may be direct where the antibody is labeled or indirect by the subsequent addition of a labeled moiety that specifically binds to the antibody as described above.

Immunoassays in the competitive binding format can be used for cross-reactivity determinations. For example, the protein encoded by the sequences described herein can be immobilized on a solid support. Proteins are added to the assay which compete with the binding of the antisera to the immobilized antigen. The ability of the above proteins to compete with the binding of the antisera to the immobilized protein is compared to that of the protein encoded by any of the sequences described herein. The percent cross-reactivity for the above proteins is calculated, using standard calculations. Those antisera with less than 10% cross-reactivity with each of the proteins listed above are selected and pooled. The cross-reacting antibodies are optionally removed from the pooled antisera by immunoabsorption with the considered proteins, *e.g.,* distantly related homologs.

The immunoabsorbed and pooled antisera are then used in a competitive binding immunoassay as described above to compare a second protein, thought to be perhaps a protein of the present invention, to the immunogen protein. In order to make this comparison, the two proteins are each assayed at a wide range of concentrations and the amount of each protein required to inhibit 50% of the binding of the antisera to the immobilized protein is determined. If the amount of the second protein required is less than 10 times the amount of the protein partially encoded by a sequence herein that is required, then the second protein is said to specifically bind to an antibody generated to an immunogen consisting of the target protein.

3. Other Assay Formats

In a particularly preferred embodiment, Western blot (immunoblot) analysis is used to detect and quantify the presence of a G protein-coupled receptor of the invention in the sample. The technique generally comprises separating sample proteins by gel electrophoresis on the basis of molecular weight, transferring the separated proteins to a suitable solid support (such as, *e.g.*, a nitrocellulose filter, a nylon filter, or a derivatized nylon filter) and incubating the sample with the antibodies that specifically bind the protein of interest. For example, the anti-GPCR antibodies specifically bind to

the G protein-coupled receptor on the solid support. These antibodies may be directly labeled or alternatively may be subsequently detected using labeled antibodies (*e.g.*, labeled sheep anti-mouse antibodies) that specifically bind to the antibodies against the protein of interest.

Other assay formats include liposome immunoassays (LIA), which use liposomes designed to bind specific molecules (*e.g.*, antibodies) and release encapsulated reagents or markers. The released chemicals are then detected according to standard techniques (*see*, Monroe *et al.*, *Amer. Clin. Prod. Rev.* 5:34-41 (1986)).

4. Reduction of Non-Specific Binding

One of skill in the art will appreciate that it is often desirable to use non-specific binding in immunoassays. Particularly, where the assay involves an antigen or antibody immobilized on a solid substrate it is desirable to minimize the amount of non-specific binding to the substrate. Means of reducing such non-specific binding are well-known to those of skill in the art. Typically, this involves coating the substrate with a proteinaceous composition. In particular, protein compositions, such as bovine serum albumin (BSA), nonfat powdered milk and gelatin, are widely used.

5. Labels

The particular label or detectable group used in the assay is not a critical aspect of the invention, as long as it does not significantly interfere with the specific binding of the antibody used in the assay. The detectable group can be any material having a detectable physical or chemical property. Such detectable labels have been well-developed in the field of immunoassays and, in general, most labels useful in such methods can be applied to the present invention. Thus, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include magnetic beads (*e.g.*, Dynabeads™), fluorescent dyes (*e.g.*, fluorescein isothiocyanate, Texas red, rhodamine, and the like), radiolabels (*e.g.*, ^{3}H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P), enzymes (*e.g.*, horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (*e.g.*, polystyrene, polypropylene, latex, *etc.*) beads.

The label may be coupled directly or indirectly to the desired component of the assay according to methods well-known in the art. As indicated above, a wide variety of labels may be used, with the choice of label depending on the sensitivity

required, the ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

Non-radioactive labels are often attached by indirect means. The molecules can also be conjugated directly to signal generating compounds, *e.g.*, by conjugation with an enzyme or fluorescent compound. A variety of enzymes and fluorescent compounds can be used with the methods of the present invention and are well-known to those of skill in the art (for a review of various labeling or signal producing systems which may be used, *see, e.g.*, U.S. Patent No. 4,391,904).

Means of detecting labels are well-known to those of skill in the art. Thus, for example, where the label is a radioactive label, means for detection include a scintillation counter or photographic film as in autoradiography. Where the label is a fluorescent label, it may be detected by exciting the fluorochrome with the appropriate wavelength of light and detecting the resulting fluorescence. The fluorescence may be detected visually, by means of photographic film, by the use of electronic detectors such as charge coupled devices (CCDs) or photomultipliers and the like. Similarly, enzymatic labels may be detected by providing the appropriate substrates for the enzyme and detecting the resulting reaction product. Finally simple colorimetric labels may be detected directly by observing the color associated with the label. Thus, in various dipstick assays, conjugated gold often appears pink, while various conjugated beads appear the color of the bead.

Some assay formats do not require the use of labeled components. For instance, agglutination assays can be used to detect the presence of the target antibodies. In this case, antigen-coated particles are agglutinated by samples comprising the target antibodies. In this format, none of the components need to be labeled and the presence of the target antibody is detected by simple visual inspection.

VII. SCREENING FOR MODULATORS OF THE GPCRs OF THE INVENTION

The invention also provides methods for identifying compounds that modulate signaling mediated by the G protein-coupled receptors of the invention. These compounds include both those that modulate the expression and those that modulate the activity of the G protein-coupled receptors of the invention. Furthermore, these compounds may modulate the expression and/or activity of one or of various G protein-coupled receptors of the invention, and optionally of all the G protein-coupled receptors

of the invention. In addition, the identified compounds can also modulate, *e.g.*, the development of Alzheimer's disease, rheumatoid arthritis, osteoarthritis, osteoporosis, amyotrophic lateral sclerosis, multiple sclerosis and atherosclerosis, asthma, depression, epilepsy, schizophrenia, Parkinson's disease, sarcomas such as, chondrosarcoma, Ewing's sarcoma, and osteosarcoma, carcinomas such as, basal cell carcinoma, breast carcinoma, embryonal carcinoma, ovarian carcinoma, renal cell carcinoma, lung adenocarcinoma, lung small cell carcinoma, pancreatic carcinoma, prostate carcinoma, transitional carcinoma of the bladder, squamous cell carcinoma, and thyroid carcinoma, psoriasis, cardiomyopathy, Crohn's disease, Duchenne muscular dystrophy, glioblastoma multiform, Hodgkin's disease, lymphoma, macular degeneration, malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, seminoma, tuberculosis, tonsil, ulcerative colitis, learning and memory processes, reproduction and sex behavior, feeding behavior, fat metabolism and body adiposity, neurotransmitter release, pain perception, depression, hormone release, cardiovascular actions, or any other disease or disorder involving GPCR-mediated signaling.

A. Screening for Modulators of the G Protein-Coupled Receptors

The present invention provides methods for identifying compounds that increase or decrease the expression level or the activity of one or more G protein-coupled receptors of interest. Compounds that are identified as modulators of the expression or activity of one or more G protein-coupled receptors of the invention using the methods described herein find use both *in vitro* and *in vivo*. For example, one can treat cell cultures with the modulators in experiments designed to determine the mechanisms by which GPCR-mediated signaling is regulated. Compounds that modulate the activity of the G protein-coupled receptors are useful for studying, for example, the mechanisms that lead to depression, Alzheimer's disease, specific sarcomas and carcinomas, other cancers such as lymphomas and melanomas, psoriasis, cardiomyopathies, *etc.* Compounds that modulate the activity of the galanin receptor are useful for studying, for example, the mechanisms that lead to growth hormone release, depression or fat accumulation, neurotransmitter or insulin release.

The methods for isolating compounds that modulate the expression of the G protein-coupled receptors of the invention typically involve culturing a cell in the presence of a potential modulator to form a first cell culture. RNA (or cDNA) from the first cell culture is contacted with one or more probes, each probe comprising a

polynucleotide sequence encoding a G protein-coupled receptor of the invention (*e.g.*, a nucleotide sequence selected from the group of sequences set forth in Table 1). The amount of the probe(s) which hybridizes to the RNA (or cDNA) from the first cell culture is determined. Typically, one determines whether the amount of the probe(s) which hybridizes to the RNA (or cDNA) is increased or decreased relative to the amount of the probe(s) which hybridizes to RNA (or cDNA) from a second cell culture grown in the absence of the modulator.

The G protein-coupled receptors of the invention and their alleles and polymorphic variants mediate signaling in different pathways involving a variety of ligands. The activity of G protein-coupled receptor polypeptides can be assessed using a variety of *in vitro* and *in vivo* assays to determine functional, chemical, and physical effects, *e.g.*, measuring ligand binding (*e.g.*, radioactive ligand binding), second messengers (*e.g.*, cAMP, cGMP, IP_3, DAG, or Ca^{2+}), ion flux, phosphorylation levels, transcription levels, neurotransmitter levels, and the like. Furthermore, such assays can be used to test for inhibitors and activators of the G protein-coupled receptors of the invention. Modulators can also be genetically altered versions of the present G protein-coupled receptors. Such modulators of GPCR-mediated signaling activity are useful for treating a variety of diseases and disorders described herein. For a general review of GPCR signal transduction and methods of assaying signal transduction, *see, e.g., Methods in Enzymology* vols. 237 and 238 (1994) and volume 96 (1983); Bourne *et al., Nature* 10:349:117-27 (1991); Bourne *et al.*, Nature 348:125-32 (1990); Pitcher *et al., Annu. Rev. Biochem.* 67:653-92 (1998).

The G protein-coupled receptors of the assay will typically be polypeptides having identity with polypeptides encoded by a nucleic acid molecule having a nucleotide sequence selected from the sequences set forth in Table 1, or conservatively modified variants thereof.

Generally, the amino acid sequence identity will be at least 70%, 75%, 80%, 85%, 90%, 95% or more identity and further will not be identical to the sequences for known GPCRs (for sequences of identified GPCRs, *see, e.g.*, http://www.gcrdb.uthscsa.edu; http://www.ncbi.nlm.nih.gov; and http://www.expasy.ch/sprot/sprot.top.html). With regard to galanin receptors, the amino acid sequences of the invention will not be identical to the sequences for GALR1, GALR2 or GALR3 (*see, e.g.*, SwissProt accession numbers P47211, O43603, and O60755 for the sequences of the human GALR1, GALR2 and GALR3, respectively).

Optionally, the polypeptide(s) of the assays will comprise a domain of a G protein-coupled receptor, such as an extracellular domain, transmembrane region, transmembrane domain, cytoplasmic domain, ligand binding domain, subunit association domain, active site, and the like. The polypeptides of the present invention may also be polypeptides comprising a region of 15 amino acids or more, optionally 30 amino acids or more, having at least 80%, preferably at least 85%, and most preferably 90% or more, identity with a region of 15 amino acids or more, optionally 30 amino acids or more, from a polypeptide encoded by a nucleic acid molecule having a nucleotide sequence selected from the group consisting of the sequences set forth in Table 1, and having substantially the same biological activity. Either the G protein-coupled receptor protein or a domain thereof can be covalently linked to a heterologous protein to create a chimeric protein used in the assays described herein.

Modulators of the activity of G protein-coupled receptors are tested using G protein-coupled receptors polypeptides as described above, either recombinant or naturally occurring. The proteins can be isolated, expressed in a cell, expressed in a membrane derived from a cell, expressed in tissue or in an animal, either recombinant or naturally occurring. For example, neurons, transformed cells, or membranes can be used. Modulation is tested using one of the *in vitro* or *in vivo* assays described herein. G protein-mediated signaling can also be examined *in vitro* with soluble or solid state reactions, using a full-length G protein-coupled receptor or a chimeric molecule such as an extracellular domain or transmembrane region, or combination thereof, of a G protein-coupled receptor covalently linked to a heterologous signal transduction domain, or a heterologous extracellular domain and/or transmembrane region covalently linked to the transmembrane and/or cytoplasmic domain of a G protein-coupled receptor. Furthermore, ligand-binding domains of the protein of interest can be used *in vitro* in soluble or solid state reactions to assay for ligand binding. In numerous embodiments, a chimeric receptor will be made that comprises all or part of a G protein-coupled receptor polypeptide as well as an additional sequence that facilitates the localization of the G protein-coupled receptor to the membrane.

Ligand binding to a G protein-coupled receptor, a domain thereof, or a chimeric protein can be tested in solution, in a bilayer membrane, attached to a solid phase, in a lipid monolayer, or in vesicles. Binding of a modulator can be tested using, *e.g.*, changes in spectroscopic characteristics (*e.g.*, fluorescence, absorbance, refractive index) hydrodynamic (*e.g.*, shape), chromatographic, or solubility properties.

G protein-coupled receptor-G protein interactions can also be examined. For example, binding of the G protein to the receptor or its release from the receptor can be examined. For example, in the absence of GTP, an activator will lead to the formation of a tight complex of a G protein (all three subunits) with the receptor. This complex can be detected in a variety of ways. Such an assay can be modified to search for inhibitors, *e.g.*, by adding an activator to the G protein-coupled receptor and G protein in the absence of GTP, which form a tight complex, and then screen for inhibitors by looking at dissociation of the G protein-coupled receptor-G protein complex. In the presence of GTP, release of the alpha subunit of the G protein from the other two G protein subunits serves as a criterion of activation.

In some embodiments, G protein-coupled receptors-ligand interactions are monitored as a function of G protein-coupled receptors activation.

An activated or inhibited G protein will in turn alter the properties of target enzymes, channels, and other effector proteins. Target enzymes and effector proteins for G protein-coupled receptors that can be used in the context of the present invention are known to those of skill in the art.

In some embodiments, a G protein-coupled receptor polypeptide is expressed in a eukaryotic cell as a chimeric receptor with a heterologous, chaperone sequence that facilitates its maturation and targeting through the secretory pathway. Chimeric G protein-coupled receptors can be expressed in any eukaryotic cell, such as HEK-293 cells. Preferably, the cells comprise a functional G protein that is capable of coupling the chimeric receptor to an intracellular signaling pathway or to a signaling protein. Activation of such chimeric receptors in such cells can be detected using any standard method, such as by detecting changes in intracellular calcium by detecting FURA-2 dependent fluorescence in the cell.

In addition, activated G protein-coupled receptors become substrates for kinases. Phosphorylation of the G protein-coupled receptors of the invention can thus also be measured as a means to detect activation of the receptors. Phosphorylation may be detected by assaying the transfer of ^{32}P from gamma-labeled GTP to the receptor with a scintillation counter.

Samples or assays that are treated with a potential G protein-coupled receptor inhibitor or activator are compared to control samples without the test compound, to examine the extent of modulation. Such assays may be carried out in the presence of ligand, and modulation of the ligand-dependent activation is monitored.

Control samples (untreated with activators or inhibitors) are assigned a relative G protein-coupled receptor activity value of 100. Inhibition of a G protein-coupled receptor protein is achieved when the G protein-coupled receptor activity value relative to the control is about 90%, optionally 50%, optionally 25-0%. Activation of a G protein-coupled receptor protein is achieved when the G protein-coupled receptor activity value relative to the control is 110%, optionally 150%, 200-500%, or 1000-2000% or more.

Changes in ion flux may be assessed by determining changes in polarization (*i.e.*, electrical potential) of the cell or membrane expressing a G protein-coupled receptor of interest. One means to determine changes in cellular polarization is by measuring changes in current (thereby measuring changes in polarization) with voltage-clamp and patch-clamp techniques, *e.g.*, the "cell-attached" mode, the "inside-out" mode, and the "whole cell" mode (*see, e.g.*, Ackerman *et al., New Engl. J. Med.* 336:1575-1595 (1997)). Whole cell currents are conveniently determined using the standard methodology (*see, e.g.*, Hamil *et al., PFlugers. Archiv.* 391:85 (1981). Other known assays include: radiolabeled ion flux assays and fluorescence assays using voltage-sensitive dyes (*see, e.g.*, Vestergarrd-Bogind *et al., J. Membrane Biol.* 88:67-75 (1988); Gonzales & Tsien, *Chem. Biol.* 4:269-277 (1997); Daniel *et al., J. Pharmacol. Meth.* 25:185-193 (1991); Holevinsky *et al., J. Membrane Biology* 137:59-70 (1994)). Generally, the compounds to be tested are present in the range from 1 pM to 100 mM.

The effects of the test compounds upon the function of the polypeptides can be measured by examining any of the parameters described above, and other parameters known to those of skill in the art. Any suitable physiological change that affects G protein-coupled receptor activity can be used to assess the influence of a test compound on the G protein-coupled receptors of this invention. When the functional consequences are determined using intact cells or animals, one can also measure a variety of effects such as transmitter release, hormone release, transcriptional changes to both known and uncharacterized genetic markers, changes in cell metabolism such as cell growth or pH changes, and changes in intracellular second messengers such as Ca^{2+}, IP3, cGMP, or cAMP.

Preferred assays for G protein-coupled receptors include cells that are loaded with ion or voltage sensitive dyes to report receptor activity. Assays for determining activity of such receptors can also use known agonists and antagonists for other G protein-coupled receptors as negative or positive controls to assess activity of tested compounds. In assays for identifying modulatory compounds (*e.g.*, agonists,

antagonists), changes in the level of ions in the cytoplasm or membrane voltage will be monitored using an ion sensitive or membrane voltage fluorescent indicator, respectively. Among the ion-sensitive indicators and voltage probes that may be employed are those disclosed in the Molecular Probes 1997 Catalog. For G protein-coupled receptors, promiscuous G proteins can be used in the assay of choice (Wilkie *et al.*, *Proc. Natl. Acad. Sci. USA* 88:10049-10053 (1991)). Such promiscuous G proteins allow coupling of a wide range of receptors.

Other assays to determine the activity of G protein-coupled receptors, can involve measuring changes in the level of intracellular cyclic nucleotides, *e.g.*, cAMP or cGMP, that occur due to the activation or inhibition of enzymes such as adenylate cyclase upon activation of the receptor.

In one embodiment, the changes in intracellular cAMP or cGMP can be measured using immunoassays. The method described in Offermanns & Simon, *J. Biol. Chem.* 270:15175-15180 (1995) may be used to determine the level of cAMP. Also, the method described in Felley-Bosco *et al.*, *Am. J. Resp. Cell and Mol. Biol.* 11:159-164 (1994) may be used to determine the level of cGMP. Further, an assay kit for measuring cAMP and/or cGMP is described in U.S. Patent No. 4,115,538.

In another embodiment, transcription levels can be measured to assess the effects of a test compound on signal transduction. A host cell containing a G protein-coupled receptor of interest is contacted with a test compound for a sufficient time to effect any interactions, and then the level of gene expression is measured. The amount of time to effect such interactions may be empirically determined, such as by running a time course and measuring the level of transcription as a function of time. The amount of transcription may be measured by using any method known to those of skill in the art to be suitable. For example, mRNA expression of the protein of interest may be detected using northern blots or their polypeptide products may be identified using immunoassays. Alternatively, transcription based assays using reporter gene may be used as described in U.S. Patent No. 5,436,128. The reporter genes can be, *e.g.*, chloramphenicol acetyltransferase, luciferase, β-galactosidase and alkaline phosphatase. Furthermore, the protein of interest can be used as an indirect reporter via attachment to a second reporter such as green fluorescent protein (*see, e.g.*, Mistili and Spector, *Nature Biotechnology* 15:961-964 (1997)). The amount of transcription is then compared to the amount of transcription in either the same cell in the absence of the test compound, or it may be

compared with the amount of transcription in a substantially identical cell that lacks the protein of interest. A substantially identical cell may be derived from the same cells from which the recombinant cell was prepared but which had not been modified by introduction of heterologous DNA. Any difference in the amount of transcription indicates that the test compound has in some manner altered the activity of the protein of interest.

Any other method that allows to determine the effect of a compounds on the activity of a G protein-coupled receptor of interest can also be used in the context of the present invention (for articles disclosing methods for determining the activity of G protein-coupled receptors, *see, e.g.*, Fisone *et al.*, *Brain Res.* 568:279-84 (1991); Ogren *et al.*, *Ann. NY Acad. Sci.* 863:342-63 (1998); Wang *et al.*, *Neuropeptides* 33:197-205 (1999)).

B. Modulators of the Activity of the G Protein-Coupled Receptors of the Invention

The compounds tested as modulators of the G protein-coupled receptors of the invention can be any small chemical compound, or a biological entity, such as a protein, sugar, nucleic acid or lipid. Alternatively, modulators can be genetically altered versions of a G protein-coupled receptor gene. Typically, test compounds will be small chemical molecules and peptides. Essentially any chemical compound can be used as a potential modulator or ligand in the assays of the invention, although most often compounds that can be dissolved in aqueous or organic (especially DMSO-based) solutions are used. The assays are designed to screen large chemical libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (*e.g.*, in microtiter formats on microtiter plates in robotic assays). It will be appreciated that there are many suppliers of chemical compounds, including Sigma (St. Louis, MO), Aldrich (St. Louis, MO), Sigma-Aldrich (St. Louis, MO), Fluka Chemika-Biochemica Analytika (Buchs, Switzerland) and the like.

In one preferred embodiment, high throughput screening methods involve providing a combinatorial chemical or peptide library containing a large number of potential therapeutic compounds (potential modulator or ligand compounds). Such "combinatorial chemical libraries" or "ligand libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus

identified can serve as conventional "lead compounds" or can themselves be used as potential or actual therapeutics.

A combinatorial chemical library is a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis, by combining a number of chemical "building blocks" such as reagents. For example, a linear combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (amino acids) in every possible way for a given compound length (*i.e.*, the number of amino acids in a polypeptide compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks.

Preparation and screening of combinatorial chemical libraries is well-known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (*see, e.g.*, U.S. Patent No. 5,010,175; Furka, *Int. J. Pept. Prot. Res*. 37:487-493 (1991); and Houghton *et al.*, *Nature* 354:84-88 (1991)). Other chemistries for generating chemical diversity libraries can also be used. Such chemistries include, but are not limited to, peptoids (*e.g.*, PCT Publication No. WO 91/19735), encoded peptides (*e.g.*, PCT Publication WO 93/20242), random bio-oligomers (*e.g.*, PCT Publication No. WO 92/00091), benzodiazepines (*e.g.*, U.S. Patent No. 5,288,514), diversomers such as hydantoins, benzodiazepines and dipeptides (Hobbs *et al.*, *Proc. Nat. Acad. Sci. USA* 90:6909-6913 (1993)), vinylogous polypeptides (Hagihara *et al.*, *J. Amer. Chem. Soc.* 114:6568 (1992)), nonpeptidal peptidomimetics with glucose scaffolding (Hirschmann *et al.*, *J. Amer. Chem. Soc.* 114:9217-9218 (1992)), analogous organic syntheses of small compound libraries (Chen *et al.*, *J. Amer. Chem. Soc.* 116:2661 (1994)), oligocarbamates (Cho *et al.*, *Science* 261:1303 (1993)), and/or peptidyl phosphonates (Campbell *et al.*, *J. Org. Chem.* 59:658 (1994)), nucleic acid libraries (*see* Ausubel *et al.*, Berger *et al.*, and Sambrook *et al.*, all *supra*), peptide nucleic acid libraries (*see, e.g.*, U.S. Patent No. 5,539,083), antibody libraries (*see, e.g.*, Vaughn *et al.*, *Nature Biotechnology*, 14(3):309-314 (1996) and PCT/US96/10287), carbohydrate libraries (*see, e.g.*, Liang *et al.*, *Science*, 274:1520-1522 (1996) and U.S. Patent No. 5,593,853), small organic molecule libraries (*see, e.g.*, benzodiazepines, Baum C&EN, Jan 18, page 33 (1993); isoprenoids, U.S. Patent No. 5,569,588; thiazolidinones and metathiazanones, U.S. Patent No. 5,549,974; pyrrolidines, U.S. Patent Nos. 5,525,735 and 5,519,134; morpholino compounds, U.S. Patent No. 5,506,337; benzodiazepines, 5,288,514, and the like), *etc.*

Devices for the preparation of combinatorial libraries are commercially available (*see, e.g.,* 357 MPS, 390 MPS, Advanced Chem Tech, Louisville KY, Symphony, Rainin, Woburn, MA, 433A Applied Biosystems, Foster City, CA, 9050 Plus, Millipore, Bedford, MA). In addition, numerous combinatorial libraries are themselves commercially available (*see, e.g.,* ComGenex, Princeton, N.J., Tripos, Inc., St. Louis, MO, 3D Pharmaceuticals, Exton, PA, Martek Biosciences, Columbia, MD, *etc.*).

C. Solid State and Soluble High Throughput Assays

In one embodiment, the invention provides soluble assays using molecules such as a domain, such as a ligand binding domain, an extracellular domain, a transmembrane domain (*e.g.*, one comprising seven transmembrane regions and cytosolic loops), the transmembrane domain and a cytoplasmic domain, an active site, a subunit association region, *etc.*, a domain that is covalently linked to a heterologous protein to create a chimeric molecule, a G protein-coupled receptor, or a cell or tissue expressing a G protein-coupled receptor, either naturally occurring or recombinant. In another embodiment, the invention provides solid phase based *in vitro* assays in a high throughput format, where the domain, chimeric molecule, G protein-coupled receptor, or cell or tissue expressing the G protein-coupled receptor is attached to a solid phase substrate.

In the high throughput assays of the invention, it is possible to screen up to several thousand different modulators or ligands in a single day. In particular, each well of a microtiter plate can be used to run a separate assay against a selected potential modulator, or, if concentration or incubation time effects are to be observed, every 5-10 wells can test a single modulator. Thus, a single standard microtiter plate can assay about 100 (*e.g.*, 96) modulators. If 1536 well plates are used, then a single plate can easily assay from about 100 to about 1500 different compounds. It is possible to assay several different plates per day. Assay screens for up to about 6,000-20,000 different compounds are possible using the integrated systems of the invention. More recently, microfluidic approaches to reagent manipulation have been developed.

The molecule of interest can be bound to the solid state component, directly or indirectly, via covalent or non covalent linkage, *e.g.*, via a tag. The tag can be any of a variety of components. In general, a molecule which binds the tag (a tag binder) is fixed to a solid support, and the tagged molecule of interest (*e.g.*, the G protein-coupled receptor of interest) is attached to the solid support by interaction of the tag and the tag binder.

A number of tags and tag binders can be used, based upon known molecular interactions well described in the literature. For example, where a tag has a natural binder, for example, biotin, protein A, or protein G, it can be used in conjunction with appropriate tag binders (avidin, streptavidin, neutravidin, the Fc region of an immunoglobulin, *etc.*) Antibodies to molecules with natural binders such as biotin are also widely available and appropriate tag binders (*see,* SIGMA Immunochemicals 1998 catalogue SIGMA, St. Louis MO).

Similarly, any haptenic or antigenic compound can be used in combination with an appropriate antibody to form a tag/tag binder pair. Thousands of specific antibodies are commercially available and many additional antibodies are described in the literature. For example, in one common configuration, the tag is a first antibody and the tag binder is a second antibody which recognizes the first antibody. In addition to antibody-antigen interactions, receptor-ligand interactions are also appropriate as tag and tag-binder pairs, such as agonists and antagonists of cell membrane receptors (*e.g.,* cell receptor-ligand interactions such as transferrin, c-kit, viral receptor ligands, cytokine receptors, chemokine receptors, interleukin receptors, immunoglobulin receptors and antibodies, the cadherin family, the integrin family, the selectin family, and the like; *see, e.g.,* Pigott and Power, *The Adhesion Molecule Facts Book I* (1993)). Similarly, toxins and venoms, viral epitopes, hormones (*e.g.*, opiates, steroids, *etc.*), intracellular receptors (*e.g.*, which mediate the effects of various small ligands, including steroids, thyroid hormone, retinoids and vitamin D; peptides), drugs, lectins, sugars, nucleic acids (both linear and cyclic polymer configurations), oligosaccharides, proteins, phospholipids and antibodies can all interact with various cell receptors.

Synthetic polymers, such as polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneimines, polyarylene sulfides, polysiloxanes, polyimides, and polyacetates can also form an appropriate tag or tag binder. Many other tag/tag binder pairs are also useful in assay systems described herein, as would be apparent to one of skill upon review of this disclosure.

Common linkers such as peptides, polyethers, and the like can also serve as tags, and include polypeptide sequences, such as poly gly sequences of between about 5 and 200 amino acids. Such flexible linkers are known to those of skill in the art. For example, poly(ethelyne glycol) linkers are available from Shearwater Polymers, Inc. Huntsville, Alabama. These linkers optionally have amide linkages, sulfhydryl linkages, or heterofunctional linkages.

Tag binders are fixed to solid substrates using any of a variety of methods currently available. Solid substrates are commonly derivatized or functionalized by exposing all or a portion of the substrate to a chemical reagent which fixes a chemical group to the surface which is reactive with a portion of the tag binder. For example, groups which are suitable for attachment to a longer chain portion would include amines, hydroxyl, thiol, and carboxyl groups. Aminoalkylsilanes and hydroxyalkylsilanes can be used to functionalize a variety of surfaces, such as glass surfaces. The construction of such solid phase biopolymer arrays is well described in the literature (*see, e.g.*, Merrifield, *J. Am. Chem. Soc.* 85:2149-2154 (1963) (describing solid phase synthesis of, *e.g.*, peptides); Geysen *et al., J. Immun. Meth.* 102:259-274 (1987) (describing synthesis of solid phase components on pins); Frank and Doring, *Tetrahedron* 44:60316040 (1988) (describing synthesis of various peptide sequences on cellulose disks); Fodor *et al., Science* 251:767-777 (1991); Sheldon *et al., Clinical Chemistry* 39(4):718-719 (1993); and Kozal *et al., Nature Medicine* 2(7):753759 (1996) (all describing arrays of biopolymers fixed to solid substrates). Non-chemical approaches for fixing tag binders to substrates include other common methods, such as heat, cross-linking by UV radiation, and the like.

The invention provides *in vitro* assays for identifying, in a high throughput format, compounds that can modulate the expression or activity of the G protein-coupled receptors of the invention. Control reactions that measure the G protein-coupled receptor activity of the cell in a reaction that does not include a potential modulator are optional, as the assays are highly uniform. Such optional control reactions are appropriate and increase the reliability of the assay. Accordingly, in a preferred embodiment, the methods of the invention include such a control reaction. For each of the assay formats described, "no modulator" control reactions which do not include a modulator provide a background level of binding activity.

In some assays it will be desirable to have positive controls to ensure that the components of the assays are working properly. At least two types of positive controls are appropriate. First, a known activator of the G protein-coupled receptors of the invention can be incubated with one sample of the assay, and the resulting increase in signal resulting from an increased expression level or activity of a G protein-coupled receptor determined according to the methods herein. Second, a known inhibitor of the G protein-coupled receptors of the invention can be added, and the resulting decrease in signal for the expression or activity of a G protein-coupled receptor similarly detected. It

will be appreciated that modulators can also be combined with activators or inhibitors to find modulators which inhibit the increase or decrease that is otherwise caused by the presence of the known modulator of the G protein-coupled receptor.

D. Computer-Based Assays

Yet another assay for compounds that modulate the activity of G protein-coupled receptors involves computer assisted drug design, in which a computer system is used to generate a three-dimensional structure of a G protein-coupled receptor based on the structural information encoded by its amino acid sequence. The input amino acid sequence interacts directly and actively with a pre-established algorithm in a computer program to yield secondary, tertiary, and quaternary structural models of the protein. The models of the protein structure are then examined to identify regions of the structure that have the ability to bind, *e.g.*, ligands. These regions are then used to identify ligands that bind to the protein.

The three-dimensional structural model of the protein is generated by entering protein amino acid sequences of at least 10 amino acid residues (or corresponding nucleic acid sequences encoding a G protein-coupled receptor) into the computer system. The nucleotide sequence encoding the GPCR can be any sequence encoding a polypeptide having at least 30%, optionally at least 40%, 50%, 60%, 70%, 80%, 90% or more identity with a polypeptide encoded by a nucleic acid molecule having a sequence selected from the group consisting of the sequences set forth in Table 1, and conservatively modified versions thereof. The amino acid sequences encoded by the nucleic acid sequences provided herein represent the primary sequences or subsequences of the proteins, which encode the structural information of the proteins. At least 10 residues of an amino acid sequence (or a nucleotide sequence encoding 10 amino acids) are entered into the computer system from computer keyboards, computer readable substrates that include, but are not limited to, electronic storage media (*e.g.*, magnetic diskettes, tapes, cartridges, and chips), optical media (*e.g.*, CD ROM), information distributed by internet sites, and by RAM. The three-dimensional structural model of the protein is then generated by the interaction of the amino acid sequence and the computer system, using software known to those of skill in the art.

The amino acid sequence represents a primary structure that encodes the information necessary to form the secondary, tertiary and quaternary structures of the protein of interest. The software looks at certain parameters encoded by the primary

sequence to generate the structural model. These parameters are referred to as "energy terms" and primarily include electrostatic potentials, hydrophobic potentials, solvent accessible surfaces, and hydrogen bonding. Secondary energy terms include van der Waals potentials. Biological molecules form the structures that minimize the energy terms in a cumulative fashion. The computer program uses these terms encoded by the primary structure or amino acid sequence to create the secondary structural model.

The tertiary structure of the protein encoded by the secondary structure is then formed on the basis of the energy terms of the secondary structure. The user at this point can enter additional variables such as whether the protein is membrane bound or soluble, its location in the body, and its cellular location, *e.g.*, cytoplasmic, surface, or nuclear. These variables along with the energy terms of the secondary structure are used to form the model of the tertiary structure. In modeling the tertiary structure, the computer program matches hydrophobic faces of secondary structure with like, and hydrophilic faces of secondary structure with like.

Once the structure has been generated, potential ligand-binding regions are identified by the computer system. Three-dimensional structures for potential ligands are generated by entering amino acid or nucleotide sequences or chemical formulas of compounds, as described above. The three-dimensional structure of the potential ligand is then compared to that of the G protein-coupled receptor to identify ligands that bind to the protein. Binding affinity between the protein and ligands is determined using energy terms to determine which ligands have an enhanced probability of binding to the protein.

Computer systems are also used to screen for mutations, polymorphic variants, alleles and interspecies homologs of genes encoding the G protein-coupled receptors of the invention. Such mutations can be associated with disease states or genetic traits. As described above, GeneChip™ and related technology can also be used to screen for mutations, polymorphic variants, alleles and interspecies homologs. Once the variants are identified, diagnostic assays can be used to identify patients having such mutated genes. Identification of the mutated G protein-coupled receptor genes involves receiving input of a first amino acid sequence of a G protein-coupled receptor (or of a first nucleic acid sequence encoding a GPCR of the invention), *e.g.*, any amino acid sequence having at least 30%, optionally at least 40%, 50%, 60%, 70%, 80%, 90% or more identity with a polypeptide encoded by a nucleic acid molecule having a sequence selected from the group consisting of the sequences set forth in Table 1, or conservatively

modified versions thereof, or alternatively any amino acid sequence comprising a region of 15 amino acids or more, optionally 30 amino acids or more, having at least 80%, preferably at least 85%, and most preferably 90% or more, identity with a region of 15 amino acids or more, optionally 30 amino acids or more, from a polypeptide encoded by a nucleic acid molecule having a nucleotide sequence selected from the group consisting of the sequences set forth in Table 1. The sequence is entered into the computer system as described above. The first nucleic acid or amino acid sequence is then compared to a second nucleic acid or amino acid sequence that has substantial identity to the first sequence. The second sequence is entered into the computer system in the manner described above. Once the first and second sequences are compared, nucleotide or amino acid differences between the sequences are identified. Such sequences can represent allelic differences in various G protein-coupled receptor genes, and mutations associated with disease states and genetic traits.

VIII. COMPOSITIONS, KITS AND INTEGRATED SYSTEMS

The invention provides compositions, kits and integrated systems for practicing the assays described herein using nucleic acids encoding the G protein-coupled receptors of the invention, or the G protein-coupled receptors proteins themselves, anti-G protein-coupled receptors antibodies, *etc.*

The invention provides assay compositions for use in solid phase assays; such compositions can include, for example, one or more nucleic acids encoding a G protein-coupled receptor immobilized on a solid support, and a labeling reagent. In each case, the assay compositions can also include additional reagents that are desirable for hybridization. Modulators of expression or activity of a G protein-coupled receptor of the invention can also be included in the assay compositions.

The invention also provides kits for carrying out the assays of the invention. The kits typically include a probe that comprises a polynucleotide sequence encoding a G protein-coupled receptor, and a label for detecting the presence of the probe. The kits may include several polynucleotide sequences encoding G protein-coupled receptors of the invention. Kits can include any of the compositions noted above, and optionally further include additional components such as instructions to practice a high-throughput method of assaying for an effect on expression of the genes encoding the G protein-coupled receptors of the invention, or on activity of the G protein-coupled receptors of the invention, one or more containers or compartments (*e.g.*, to hold the

probe, labels, or the like), a control modulator of the expression or activity of G protein-coupled receptors, a robotic armature for mixing kit components or the like.

The invention also provides integrated systems for high-throughput screening of potential modulators for an effect on the expression or activity of the G protein-coupled receptors of the invention. The systems typically include a robotic armature which transfers fluid from a source to a destination, a controller which controls the robotic armature, a label detector, a data storage unit which records label detection, and an assay component such as a microtiter dish comprising a well having a reaction mixture or a substrate comprising a fixed nucleic acid or immobilization moiety.

A number of robotic fluid transfer systems are available, or can easily be made from existing components. For example, a Zymate XP (Zymark Corporation; Hopkinton, MA) automated robot using a Microlab 2200 (Hamilton; Reno, NV) pipetting station can be used to transfer parallel samples to 96 well microtiter plates to set up several parallel simultaneous STAT binding assays.

Optical images viewed (and, optionally, recorded) by a camera or other recording device (*e.g.,* a photodiode and data storage device) are optionally further processed in any of the embodiments herein, *e.g.,* by digitizing the image and storing and analyzing the image on a computer. A variety of commercially available peripheral equipment and software is available for digitizing, storing and analyzing a digitized video or digitized optical image, *e.g.,* using PC (Intel x86 or Pentium chip-compatible DOS®, OS2® WINDOWS®, WINDOWS NT®, WINDOWS95® or WINDOWS98® based computers), MACINTOSH®, or UNIX® based (*e.g.,* SUN® work station) computers.

One conventional system carries light from the specimen field to a cooled charge-coupled device (CCD) camera, in common use in the art. A CCD camera includes an array of picture elements (pixels). The light from the specimen is imaged on the CCD. Particular pixels corresponding to regions of the specimen (*e.g.,* individual hybridization sites on an array of biological polymers) are sampled to obtain light intensity readings for each position. Multiple pixels are processed in parallel to increase speed. The apparatus and methods of the invention are easily used for viewing any sample, *e.g.,* by fluorescent or dark field microscopic techniques.

IX. GENE THERAPY APPLICATIONS

A variety of human diseases can be treated by therapeutic approaches that involve stably introducing a gene into a human cell such that the gene is transcribed and

the gene product is produced in the cell. Diseases amenable to treatment by this approach include inherited diseases, including those in which the defect is in a single gene. Gene therapy is also useful for treatment of acquired diseases and other conditions. For discussions on the application of gene therapy towards the treatment of genetic as well as acquired diseases, *see*, Miller, *Nature* 357:455-460 (1992); and Mulligan, *Science* 260:926-932 (1993).

In the context of the present invention, gene therapy can be used for treating a variety of disorders and/or diseases in which G protein-coupled receptor-mediated signaling has been implicated. For example, introduction by gene therapy of polynucleotides encoding a G protein-coupled receptor of the invention can be used to treat, *e.g.*, Alzheimer's disease, rheumatoid arthritis, osteoarthritis, osteoporosis, amyotrophic lateral sclerosis, multiple sclerosis and atherosclerosis, asthma, depression, epilepsy, schizophrenia, Parkinson's disease, a number of sarcomas (*e.g.*, chondrosarcoma, Ewing's sarcoma, osteosarcoma, *etc.*) and carcinomas (*e.g.*, basal cell carcinoma, breast carcinoma, embryonal carcinoma, ovarian carcinoma, renal cell carcinoma, lung adenocarcinoma, lung small cell carcinoma, pancreatic carcinoma, prostate carcinoma, transitional carcinoma of the bladder, squamous cell carcinoma, thyroid carcinoma, *etc.*), psoriasis, cardiomyopathy, Crohn's disease, Duchenne muscular dystrophy, glioblastoma multiform, Hodgkin's disease, lymphoma, macular degeneration, malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, seminoma, tuberculosis, tonsil, ulcerative colitis, *etc.* Introduction by gene therapy of polynucleotides encoding a galanin receptor of the invention can be used to treat, *e.g.*, anorexia, to induce nerve regeneration and to decrease noniception. In addition, antisense polynucleotides can also be administered using gene therapy to treat, *e.g.*, obesity, diabetes

A. Vectors for Gene Delivery

For delivery to a cell or organism, the nucleic acids of the invention can be incorporated into a vector. Examples of vectors used for such purposes include expression plasmids capable of directing the expression of the nucleic acids in the target cell. In other instances, the vector is a viral vector system wherein the nucleic acids are incorporated into a viral genome that is capable of transfecting the target cell. In a preferred embodiment, the nucleic acids can be operably linked to expression and control

sequences that can direct expression of the gene in the desired target host cells. Thus, one can achieve expression of the nucleic acid under appropriate conditions in the target cell.

B. Gene Delivery Systems

Viral vector systems useful in the expression of the nucleic acids include, for example, naturally occurring or recombinant viral vector systems. Depending upon the particular application, suitable viral vectors include replication competent, replication deficient, and conditionally replicating viral vectors. For example, viral vectors can be derived from the genome of human or bovine adenoviruses, vaccinia virus, herpes virus, adeno-associated virus, minute virus of mice (MVM), HIV, sindbis virus, and retroviruses (including, but not limited to, Rous sarcoma virus), and MoMLV. Typically, the genes of interest are inserted into such vectors to allow packaging of the gene construct, typically with accompanying viral DNA, followed by infection of a sensitive host cell and expression of the gene of interest.

As used herein, "gene delivery system" refers to any means for the delivery of a nucleic acid of the invention to a target cell. In some embodiments of the invention, nucleic acids are conjugated to a cell receptor ligand for facilitated uptake (*e.g.*, invagination of coated pits and internalization of the endosome) through an appropriate linking moiety, such as a DNA linking moiety (*see, e.g.*, Wu *et al.*, *J. Biol. Chem.* 263:14621-14624 (1988); and WO 92/06180). For example, nucleic acids can be linked through a polylysine moiety to asialo-oromucocid, which is a ligand for the asialoglycoprotein receptor of hepatocytes.

Similarly, viral envelopes used for packaging gene constructs that include the nucleic acids of the invention can be modified by the addition of receptor ligands or antibodies specific for a receptor to permit receptor-mediated endocytosis into specific cells (*see, e.g.*, WO 93/20221; WO 93/14188; and WO 94/06923). In some embodiments of the invention, the DNA constructs of the invention are linked to viral proteins, such as adenovirus particles, to facilitate endocytosis (Curiel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 88:8850-8854 (1991)). In other embodiments, molecular conjugates of the instant invention can include microtubule inhibitors (WO 94/06922), synthetic peptides mimicking influenza virus hemagglutinin (Plank *et al.*, *J. Biol. Chem.* 269:12918-12924 (1994)), and nuclear localization signals such as SV40 T antigen (WO 93/19768).

Retroviral vectors are also useful for introducing the nucleic acids of the invention into target cells or organisms. Retroviral vectors are produced by genetically

manipulating retroviruses. The viral genome of retroviruses is RNA. Upon infection, this genomic RNA is reverse transcribed into a DNA copy which is integrated into the chromosomal DNA of transduced cells with a high degree of stability and efficiency. The integrated DNA copy is referred to as a provirus and is inherited by daughter cells as is any other gene. The wild type retroviral genome and the proviral DNA have three genes, the *gag*, the *pol* and the *env* genes, which are flanked by two long terminal repeat (LTR) sequences. The *gag* gene encodes the internal structural (nucleocapsid) proteins; the *pol* gene encodes the RNA directed DNA polymerase (reverse transcriptase); and the *env* gene encodes viral envelope glycoproteins. The 5' and 3' LTRs serve to promote transcription and polyadenylation of virion RNAs. Adjacent to the 5' LTR are sequences necessary for reverse transcription of the genome (the tRNA primer binding site) and for efficient encapsulation of viral RNA into particles (the Psi site) (*see*, Mulligan, *In*: *Experimental Manipulation of Gene Expression*, Inouye (ed), 155-173 (1983); Mann *et al.*, *Cell* 33:153-159 (1983); Cone and Mulligan, *Proc. Natl. Acad. Sci. U.S.A.* 81:6349-6353 (1984)).

The design of retroviral vectors is well-known to those of ordinary skill in the art. In brief, if the sequences necessary for encapsidation (or packaging of retroviral RNA into infectious virions) are missing from the viral genome, the result is a *cis* acting defect which prevents encapsidation of genomic RNA. However, the resulting mutant is still capable of directing the synthesis of all virion proteins. Retroviral genomes from which these sequences have been deleted, as well as cell lines containing the mutant genome stably integrated into the chromosome are well-known in the art and are used to construct retroviral vectors. Preparation of retroviral vectors and their uses are described in many publications including, *e.g.*, European Patent Application EPA 0 178 220; U.S. Patent No. 4,405,712; Gilboa, *Biotechniques* 4:504-512 (1986); Mann *et al.*, *Cell* 33:153-159 (1983); Cone and Mulligan, *Proc. Natl. Acad. Sci. USA* 81:6349-6353 (1984); Eglitis *et al.*, *Biotechniques* 6:608-614 (1988); Miller *et al.*, *Biotechniques* 7:981-990 (1989); Miller (1992) *supra*; Mulligan (1993), *supra*; and WO 92/07943.

The retroviral vector particles are prepared by recombinantly inserting the desired nucleotide sequence into a retrovirus vector and packaging the vector with retroviral capsid proteins by use of a packaging cell line. The resultant retroviral vector particle is incapable of replication in the host cell but is capable of integrating into the host cell genome as a proviral sequence containing the desired nucleotide sequence. As a

result, the patient is capable of producing, for example, a G protein-coupled receptor of interest and thus restore the cells to a normal phenotype.

Packaging cell lines that are used to prepare the retroviral vector particles are typically recombinant mammalian tissue culture cell lines that produce the necessary viral structural proteins required for packaging, but which are incapable of producing infectious virions. The defective retroviral vectors that are used, on the other hand, lack these structural genes but encode the remaining proteins necessary for packaging. To prepare a packaging cell line, one can construct an infectious clone of a desired retrovirus in which the packaging site has been deleted. Cells comprising this construct will express all structural viral proteins, but the introduced DNA will be incapable of being packaged. Alternatively, packaging cell lines can be produced by transforming a cell line with one or more expression plasmids encoding the appropriate core and envelope proteins. In these cells, the *gag*, *pol*, and *env* genes can be derived from the same or different retroviruses.

A number of packaging cell lines suitable for the present invention are also available in the prior art. Examples of these cell lines include Crip, GPE86, PA317 and PG13 (*see* Miller *et al.*, *J. Virol.* 65:2220-2224 (1991)). Examples of other packaging cell lines are described in Cone and Mulligan, *Proc. Natl. Acad. Sci. USA* 81:6349-6353 (1984); Danos and Mulligan, *Proc. Natl. Acad. Sci. USA* 85:6460-6464 (1988); Eglitis *et al.* (1988), *supra*; and Miller (1990), *supra*.

Packaging cell lines capable of producing retroviral vector particles with chimeric envelope proteins may be used. Alternatively, amphotropic or xenotropic envelope proteins, such as those produced by PA317 and GPX packaging cell lines may be used to package the retroviral vectors.

In some embodiments of the invention, an antisense nucleic acid is administered which hybridizes to a gene encoding a G protein-coupled receptor of the invention or to a transcript thereof. The antisense nucleic acid can be provided as an antisense oligonucleotide (*see, e.g.*, Murayama *et al.*, *Antisense Nucleic Acid Drug Dev.* 7:109-114 (1997)). Genes encoding an antisense nucleic acid can also be provided; such genes can be introduced into cells by methods known to those of skill in the art. For example, one can introduce a gene that encodes an antisense nucleic acid in a viral vector, such as, for example, in hepatitis B virus (*see, e.g.*, Ji *et al.*, *J. Viral Hepat.* 4:167-173 (1997)), in adeno-associated virus (*see, e.g.*, Xiao *et al.*, *Brain Res.* 756:76-83 (1997)), or in other systems including, but not limited, to an HVJ (Sendai virus)-liposome gene

delivery system (*see, e.g.*, Kaneda *et al.*, *Ann. NY Acad. Sci.* 811:299-308 (1997)), a "peptide vector" (*see, e.g.*, Vidal *et al.*, *CR Acad. Sci III* 32:279-287 (1997)), as a gene in an episomal or plasmid vector (*see, e.g.*, Cooper *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 94:6450-6455 (1997), Yew *et al.*, *Hum Gene Ther.* 8:575-584 (1997)), as a gene in a peptide-DNA aggregate (*see, e.g.*, Niidome *et al.*, *J. Biol. Chem.* 272:15307-15312 (1997)), as "naked DNA" (*see, e.g.*, U.S. Patent Nos. 5,580,859 and 5,589,466), in lipidic vector systems (*see, e.g.*, Lee *et al.*, *Crit Rev Ther Drug Carrier Syst.* 14:173-206 (1997)), polymer coated liposomes (U.S. Patent Nos. 5,213,804 and 5,013,556), cationic liposomes (Epand *et al.*, U.S. Patent Nos. 5,283,185; 5,578,475; 5,279,833; and 5,334,761), gas filled microspheres (U.S. Patent No. 5,542,935), ligand-targeted encapsulated macromolecules (U.S. Patent Nos. 5,108,921; 5,521,291; 5,554,386; and 5,166,320).

C. Pharmaceutical Formulations

When used for pharmaceutical purposes, the vectors used for gene therapy are formulated in a suitable buffer, which can be any pharmaceutically acceptable buffer, such as phosphate buffered saline or sodium phosphate/sodium sulfate, Tris buffer, glycine buffer, sterile water, and other buffers known to the ordinarily skilled artisan such as those described by Good *et al.*, *Biochemistry* 5:467 (1966).

The compositions can additionally include a stabilizer, enhancer or other pharmaceutically acceptable carriers or vehicles. A pharmaceutically acceptable carrier can contain a physiologically acceptable compound that acts, for example, to stabilize the nucleic acids of the invention and any associated vector. A physiologically acceptable compound can include, for example, carbohydrates, such as glucose, sucrose or dextrans, antioxidants, such as ascorbic acid or glutathione, chelating agents, low molecular weight proteins or other stabilizers or excipients. Other physiologically acceptable compounds include wetting agents, emulsifying agents, dispersing agents or preservatives, which are particularly useful for preventing the growth or action of microorganisms. Various preservatives are well-known and include, for example, phenol and ascorbic acid. Examples of carriers, stabilizers or adjuvants can be found in Remington's *Pharmaceutical Sciences*, Mack Publishing Company, Philadelphia, PA, 17th ed. (1985).

D. Administration of Formulations

The formulations of the invention can be delivered to any tissue or organ using any delivery method known to the ordinarily skilled artisan. In some embodiments

of the invention, the nucleic acids of the invention are formulated in mucosal, topical, and/or buccal formulations, particularly mucoadhesive gel and topical gel formulations. Exemplary permeation enhancing compositions, polymer matrices, and mucoadhesive gel preparations for transdermal delivery are disclosed in, *e.g.*, U.S. Patent No. 5,346,701.

E. Methods of Treatment

The gene therapy formulations of the invention are typically administered to a cell. The cell can be provided as part of a tissue, such as an epithelial membrane, or as an isolated cell, such as in tissue culture. The cell can be provided *in vivo*, *ex vivo*, or *in vitro*.

The formulations can be introduced into the tissue of interest *in vivo* or *ex vivo* by a variety of methods. In some embodiments of the invention, the nucleic acids of the invention are introduced into cells by such methods as microinjection, calcium phosphate precipitation, liposome fusion, or biolistics. In further embodiments, the nucleic acids are taken up directly by the tissue of interest.

In some embodiments of the invention, the nucleic acids of the invention are administered *ex vivo* to cells or tissues explanted from a patient, then returned to the patient. Examples of *ex vivo* administration of therapeutic gene constructs include Nolta *et al.*, *Proc Natl. Acad. Sci. USA* 93(6):2414-9 (1996); Koc *et al.*, *Seminars in Oncology* 23 (1):46-65 (1996); Raper *et al.*, *Annals of Surgery* 223(2):116-26 (1996); Dalesandro *et al.*, *J. Thorac. Cardi. Surg.* 11(2):416-22 (1996); and Makarov *et al.*, *Proc. Natl. Acad. Sci. USA* 93(1):402-6 (1996).

X. ADMINISTRATION AND PHARMACEUTICAL COMPOSITIONS

Modulators of the G protein-coupled receptors of the present invention can be administered directly to the mammalian subject for modulation of G protein-coupled receptor signaling *in vivo*. Administration is by any of the routes normally used for introducing a modulator compound into contact with the tissue to be treated and well-known to those of skill in the art. Although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

The pharmaceutical compositions of the invention may comprise a pharmaceutically acceptable carrier. Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular

method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions of the present invention (*see, e.g.*, Remington, *Pharmaceutical Sciences*, 17^{th} ed. 1985)).

The modulators of the expression or activity of the G protein-coupled receptors of the invention, alone or in combination with other suitable components, can be made into aerosol formulations (*i.e.*, they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

Formulations suitable for administration include aqueous and non-aqueous solutions, isotonic sterile solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. In the practice of this invention, compositions can be administered, for example, orally, nasally, topically, intravenously, intraperitoneally, or intrathecally. The formulations of compounds can be presented in unit-dose or multi-dose sealed containers, such as ampoules and vials. Solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described. The modulators can also be administered as part a of prepared food or drug.

The dose administered to a patient, in the context of the present invention should be sufficient to effect a beneficial response in the subject over time. The dose will be determined by the efficacy of the particular modulators employed and the condition of the subject, as well as the body weight or surface area of the area to be treated. The size of the dose also will be determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular compound or vector in a particular subject.

In determining the effective amount of the modulator to be administered a physician may evaluate circulating plasma levels of the modulator, modulator toxicity, and the production of anti-modulator antibodies. In general, the dose equivalent of a modulator is from about 1 ng/kg to 10 mg/kg for a typical subject.

For administration, the GPCR modulators of the present invention can be administered at a rate determined by the LD-50 of the modulator, and the side-effects of the inhibitor at various concentrations, as applied to the mass and overall health of the subject. Administration can be accomplished via single or divided doses.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to one of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

Table 1 below indicates, by identification in the "LifeSpan Cluster ID" column, sequences encoding putative human G protein-coupled receptors that were identified by low-stringency protein- and DNA-based blast searches of publicly available databases. "Acc. No" indicates the accession number of the sequence in the database from which the sequence of each putative receptor was identified. The type of database from which the sequence was identified and the length of the sequence in base-pairs (bp) are indicated in the "Database type" and the "Sequence Length" columns, respectively. The sequence is shown in the "Sequence" column. The column designated "LS Cluster Name and/or Representative Sequence (SEQ ID NO) provides the name of LifeSpan's gene cluster for the sequence as well as the sequence ID of another representative sequence for the cluster, if available. These representative sequences are provided in the sequence listing following Table 1. Table 1 further shows information about the closest homolog of the sequence. The name, accession number and length of the closest homolog are shown in the "Homolog Name," "Homolog Accession No." and "Len" columns, respectively. Length is given in number of amino acids unless otherwise indicated. The table also indicates the position ("From" and "To" columns) and length ("Aligned") of the region of significant identity between the sequence of interest and its closest homolog, as well as the percent identity ("Percent") over the described region.

Table 1

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
22315 LG5261	AC006087	Genomic Clone	1237	CAATGGCATG TGTGTTCAGA GGGCGGAATC CTGGGGACAC TGTGTGAAGG AAGACAGAGA GTGGGAGTCG GAGGGTCGGA GCAGCCCCTG ACTGGCGGCA TCCGGCCTGG TGGCGTCGGT GGTGACGGCG GACCTTTCGG ATTGCGCGAG CGCCGCCCGC GTCCCGTTGG TGGCCGAGGT CCTGGCCCGG TGCGGAGTGC CCAGGCCGCG CAGGGTGTTG CGGAAGCCCT CGGCGCTAAA GTAGTACACC AGCGGGTCCA GCACGCAGTT GGCGCCGGCC AGCAGCACCA TCACCATCAG CACCCCGCGC ACGCGATCGC GGGCAGGCAC GCTGGCCGCC ACCAGCTTGC TCCGCAGCAG CCCGTAGACC GCCAGCGTGC TGTTGTAGGG CACGAAGCAC AGCAGGAAGA TGACGAGGTT AGCCAGCAGG AGGCGCACGG TCTTCCGCCG CCGCTGGCTC TGCGTGGCGT CGGGGCGCGC CAGCGTCCAG AAGACTCGGC CCGACGAGTA GACCACCGCC GCCAGGGGCA GCAGGAAGCC CAGCGCCTCG GCCAGCAGCA CGAGGGGCAG CAGCCTGCCT TTCCACAGCT CGTCGCTGAA GCTCTCGAAG CATAGGCGCA CCTCGAGGTC CCGGTAGCGG CAACGCGAGG GCCTGTGCAC GCGGGCGGCG GGCACGGCAA ACACCAGGAT GAGCGCCCAC ACGCCCAGGC AGAGCAGCCG CGCCACGCGG GGCCGCCGCA GGTGGCGCAG TCGCAGCGGG TGCACGATGG CGGCGTAGCG GTCCACGTTG ATGAGCATCA GGAAGATGCA GCTGCCGTAC ATGTTCATCT GGAAGATGGC GCCCGTCGTC TGGCACAGGA GGTCGGGGAA GGGCCAGTGG TGCAGTGCGT AGTAGGAGAG ACGAACGGGC AGCGAGAGGG TGAAGAGCAG GTCGCTGGCC GCCAGGTTAC ACATGTACAC GCTCACCACC GAGTGCACGC GCAGCGCGCG CAGGAAGACC CAGAGGGCTA GCGCGTTGAG GGGGAGCCCG GCAGCCAGCA CCAAGCTGTA GACCACCAAG TGCAGGCGGT GGGTAGGTCG GTAGTCAGGA CACGGGAGAA CAGAACTGTT GGTTGAGGAG CTGTTGGCTA ACATCGTGCC AAAGTGGGAT TGGGAGCTAG GCTGGGGATG CCATGGAGCA CACCAGAATC ATGGCATGGC ATTCACCTCC GGGGCTGGGG CCTAGAGGCT GTACAGA	GPR92 SEQ ID NO:3)	P43657	P2Y Purinoreceptor 5	344	17	303	288	37

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
30875	AF086432	GenBank	589	ATCCATGACT GCTCAAAACT TAAAAGTCCT TTGGGGGTCA AATGGCATAC GGCAGTCACC TATGTGAACA GCTGCTTGTT TGTGGCCGTG CTGGTGATTC TGATCGGATG TTACATAGCC ATATCCAGGT ACATCCACAA ATCCAGCAGG CAATTCATAA GTCAGTCAAG CCGAAAGCGA AAACATAACC AGAGCATCAG GGTTGTTGTG GCTGTGTTTT TTACCTGCTT TCTACCATAT CACTTGTGCA GAATTCCTTT TACTTTTAGT CACTTAGACA GGCTTTTAGA TGAATCTGCA CAAAAAATCC TATATTACTG CAAAGAAATT ACACTTTTCT TGTCTGCGTG TAATGTTTGC CTGGATCCAA TAATTTACTT TTTCATGTGT AGGTCATTTT CAAGAAGGCT GTTCAAAAAA TCAAATATCA GAACCAGGAG TGAAAGCATC AGATCACTGC AAAGTGTGAG AAGATCGGAA GTTCGCATAT ATTATGATTA CACTGATGTG TAGGCCTTTT ATTGTTTGTT GGAATCGATA TGTACAAAGT GTAAATAAAT GTTTCTTTTC ATTATCCTTA AAAAAAAAA	GPR 87 SEQ ID NO:4	Q15391	ORF	338 bp	172	332	160	41
191172 (46930)	AA758208	Dbest	986	TTTTAATTTA CACTGGTATG TTTATTTTCA AAAGGCTGGC AAAGGGGATC AGAAATGGGA CTTTTTACTC AAGGGTCCTG ATATGAGGGA GGCAGGGTAG TGAGAGAAGG TCACTAATTC TTTGGCCACT CAAAACATTA ACAATGTGCT ACCTTGACTC TGTAAATCAT TGATGACATC TTTAAGTTCA CCTGAAATAA CAAAAAGTTG GGTCTGTAAG GGAACCTAGT AGCCTGTGGG CATGCTTGCC AAGCACTGAA GTCCTATCTC CCCACTTCTC CTACAGATAA CGAGCTGCAC GGCCAAGAGA GTCACAATTC AGGCAACAGG AGCGACGGGC CAGGAAAGAA CACCACCCTT CACAATGAAT TTGACACAAT TGTCTTGCCG GTGCTTTATC TCATTATATT TGTGGCAAGC ATCTTGCTGA ATGGTTTAGC AGTGTGGATC TTCTTCCACA TTAGGAATAA AACCAGCTTC ATATTCTATC TCAAAGACAT AAGTGGTTGC AGACCTCATA ATGACGCTGA CATTTCCATT TCGCATAGTC CATGATGCAG GATTTGGGAC CTTGGTACTT CAGGTTTATT CTCTGCAGAT ACACTTCCAG TTTGGTTTAT GGCAACCTGT ATACTTCCAT CGTGTTTCCT GGGGCTGATA GGCATTGGAT CGCTATCTGG AAGGTGGTCA AGGCATTTTG GGGACTTCTC GGATGTACAG GCATTAACCC TCCAGAAGGG TTTCATCTGT TTTGTGTTTG GGGGGATCAT CGCCTGTTTT GTCCTTTGCC CAACAATCAT TCCGGACAAA TGGGTCAGCC ACCCCGAGGA CGATTTTTCA TGGACTGGCC CAGAACTTAA AAGGTCCTCT GGGGGGTCAA AATGGATAAC GGGCGCCACC CCATGGGAAA CCGCCTGCCT GGTTGGTGGA CGCTCATGGT GATATCAGAT CGGGATGGTA ACAAAGGCCA TATCCCGGGT AAATCC		Q15391	ORF	338 bp	24	63	40	55

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len	From	To	Aligned	Percent
54602	AA012849	Dbest	447	TCCACATCTT CACACTCCTT CAGGATCAAA AACCTAAGCC ACATGACTGG ATGAGCCGTC ACTTGGCCTT CATTCGGGTA GTGATGGTCC TCACTGTAGT GGATGTTTTG CCTCCAGATA TGCTTGAATC ACTGCATTTT GGGAATAACT TCAAATGCAA GTCCTTGATC TAAATAAACA GAGTGACGAG GGGCCTATGT ATCTATACCA CCTGTCTCCT GAGTATACAC CAGGCCAGCA TAATCAGCCT CAGCAACTTC TGGTTGGAAA GCTTTAAACA TAAATTTACA AATAACATTG TCAGTGTCCT CTTTTTTCTT TTTTGTTCCC TCAATTTGTC TTTCAGTAGT GACATAATAT TCTTCACTGT GGCTTCTTCC ATTGTGACCC AGACCAATCT ACTTAAGGTC CGCAAATACT GCTCACGTTC TCCCATGAAG TCCATCA	Phermone receptor (PHRET) SEQ ID NO:5	Q62855	Pheromone receptor VN6, rat	310	35	182	142	34
55728	AI024852	Dbest	505	TTTTTTCATT TAGATAACAT TTTATTTGTT AGAGCAAGAT TTGGCAGACT TCATTTCAAC AGCTTAAACA TCCAAAACAA CAGGGGACAT TTTTGAACAA TCTGTAATAT TCTTCTTGAA TCTTTCTAGA TAAAACACAC AGGAATAAAA AAATGAACAT CCCCTGGAAA GCATTGCTGA CTGTGAAGAG GTAAGCTGTA ACCACTGATG CGTGCACAAC ATGGAGAACC CCAAAGATCC AGGTGGTGCC GAGAAGGAAC AGAAGAGCGA GGGCTCCTCT TGCACAAGAC CTTATGTTCT CAAAGCAACT AACTTCTGGT TTCAACCCTG CAGTGTGACG AAAAACTTTG TATATGATGA CTCCAAAAGC CAAGAGATTA ACAAGAATGA TTAGGCATGC TGGTCCTATA AAACTCCAAA TAAAGTGTGT TTCGGTGCTA AGCCAACATA CTTNTGTTGT GCCATAATAT CTGTATCCTA GTGCTGCCGA AAATCCAACT ACCAC	ETL protein SEQ ID NO:6	O94867	KIAA0768 Protein [Fragment].	872	405	548	145	53
160221 (121660)	T19393	Dbest	379	GCACATTCGT CCTCCTAACT CGACTTTCTT CCTGACAATA GGCCCTGCAG TCTTTTTGTA GCGGTACTGA CGTCTTTTAT TCCATGTGTG GTTCCTTTTT TTCTTTTTCT ATAAAGGCTG TACTAATTTT CTTCATGCAA CGTTTCCTAA AGACCATGGC CAGTTTTCTA CAGAAGCTAT TTTTGACAAC CTCAAGTGGC ATTACATTTT GCAGTGAAGT AGAGGAACCT AGGGGGACTT CTTCACAAGT TAGATTTCTT GAGGATCTTC TGTTGGAAGC AGGAGAAAGT GGGGGGTGGG GGGAAGTTGT CCGAAATGCC CTCTGAATTG CCGGCTGCAG GGTCCTTGTG CTGCGCTGGT TCTTTGAAAG TCTCAGTGT	GPR27 SEQ ID NO:7	AF027955	G protein-coupled receptor (GPR27), Mus musculus	2679 bp	2548	2626	79	93

Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
AATGCTAAAA GTACTGGTTG GCTCTGTAGG ACCCTCAGAA TCAAAAGGAA ACTCCTCCAC ACTTTGTCTC TGTCTTCTCC AGGACCCATA TTTCTTGGCC ACTTTCATAA CGTAGTTTTT GAAAGATGCT CCCATAAAAA CATAAAGGAT TGGGTTGAGG CAGCTGTGAA AGAGTGCGAT GCTTTCTGTG ACTTGGATGG CGATGTCCAT GCGTTTGCTC ATGTTGCAGC TGGTGATCAG GGAGTAGATG ATGTCTATGG CTCGGCAGAA CTTGACAATG TTATAAGGCA GTTGAGTGAC AATGAAAACT ATAACGACTG TGAGCAGAAC TTTTAGGGGT CGAGATATTT TAATGTTTGG CATCTTCATG AGTGTCCTTG CTGTGATAAA GTAGCACACC CCCATAATAA GAAAGGGTAC TACAAATCCA ATGCAGATCT CTAGCATTTG AATCAATGCT TTCATTGATG TTCCTAGGTA GCGGGGGAAA ATGGGAATGC ACCTAGCATT GTCATTTACT GTATAAAAAA CCAGCTGGGG TATGCTCAGC AAGATGGCAG CCATCCAGAC ACAGAAACAG ATGATCCAGC ATGGTTTTCC CACTCCTGAT TGGCTGGGGA CTTTAGTTAC TGCCACATAT CTGTCTATGC TGATACAAGC CAGAAACTGC ATTCCAGAGA CAAAGTTTAG TGTGTACAAG GCTGAAGTTA TTTTGCACAT TATTTTCCCT AAAACCCACC CATGAACTGC ATTAACAGCC CAAAAAGGCA GAGTGAATAG AAGGAGTAAA TCTGCTACAG CCAAATTCAG GATGTACACA TCTGTTTTGG TTCTCTGTTT CTTGTAATAG GCATAAATTG CCACTACCAT GGAATTGCCT GCAAGTCCAA TGACGAAAAC TATTGTGAGG AATACAGGGA GGAAAACTTT TGCAAATTCT CTGACATCTT CTTTGATACA GATCAGTTCA TATTGACTGT AGTCATAAGT GCCATTCATT TCATTTTCCT CATAATAATA ATCTGTTGAC TGGTTCTGTT CCAAAGCCAT GGCTCCAAT	C-C chemokine receptor 11 SEQ ID NO:34	P32248	C-C Chemokine receptor type 7 precursor	378	21	156	139	48
TTTTTGAAGT TTTCATTCAT AAATGCATAG ACAATGGGAT TACAGATGGA GTTGGAAAAT CCAATAATTT GCACGATAGC AAAAATCATC TTGATTGTGA CATCATCATA TTCCTTTTCA AAATTACTGT ATTCAATCAT CATATGGACA ACATGGAATG GTGCCCAGCA CACAGCAAAG AGAGCCACCA CTGTCACCAT CATAATGACA GCTCGTTTCT TCTTCCATAA GAGGCAGGAG GAAGAGGATG ACAAGGATGA AGGTGGTGTA GATCTTCTGG TGCACAGGGC TGGTCCACTC TTCTAAGCAG CAGATGTGTT CCTTTTCATA TAGGAAGTCA TATTTGATCT CAAGTTGTTG CACGTGCCAC ATGGGTGATC CTACGATGAC TGCCACCAGC CAGACCACAC CTAGCATTGT GAAAGCCCTT CG	SEQ ID NO:8	Q9Y5X5	G protein-coupled receptor	522	260	334	68	33

LS Cluster ID: Current (Original) LG NO.	Acc. No.	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
160324	AI090920	Dbest	455	ACTTTGCCTT CCAGCTTTTT GTTGTTTTTT CTGTCCTTAC TTTTGGACTT TCTATAAGAA TCATATACTT TTTTTGCAAT AACCACATAA AACACAAGCA TTAGGATAAA AACAGTCCAG AAAATAAACT GGCATATGTT ATTTACCATT TGATGCCATT TCAGCCCCAG AGGCCCCTTT AAGGAAGCAC ACTTTTTCAC AGACGATGGT GTTGCTTCCT TGTTGCTCAA GATCATATTT GGCAGGGAGA TGAAGAACAA AAAGAACCAG ATGAAGATTG AGACCGTTTT TGCAAAAACA GGTTTTTTTA GAAAAATATT TCTCAAAGGT CTGATGATCT TGAGGAATCT GTCAAAGGCT ATGAGCGCTA ACAGCACGAT GCCCACATAC ATGGTCTCAT AAAATATCAC CGAAGAAAAA CGACACACAA AAGCTCTGAG CTGCCAGGGT GCCAG	GPR86 SEQ ID NO:9	Q15391	ORF, complete cds	338	84	230	147	43
160435	AA804531	Dbest	599	AACTGGAAGG GCAGCCGTCT GCCGCCCACG AACACCTTCT CAAGCACTTT GAGTGACCAC GGCTTGCAAG CTGGTGGCTG GCCCCCCGAG TCCCGGGCTC TGAGGCACGG CCGTCGACTT AAGCGTTGCA TCCTGTTACC TGGAGACCCT CTGAGCTCTC ACCTGCTACT TCTGCCGCTG CTTCTGCACA GAGCCCGGGC GAGGACCCCT CCAGGATGCA GGTCCCGAAC AGCACCGGCC CGGACAACGC GACGCTGCAG ATGCTGCGGA ACCCGGCGAT CGCGGTGGCC CTGCCCGTGG TGTACTCGCT GGTGGCGGCG GTCAGCATCC CGGGCAACCT CTTCTCTCTG TGGGTGCTGT GCCGGCGCAT GGGGCCCAGA TCCCCGTCGG TCATCTTCAT GATCAACCTG AGCGTCACGG ACCTGATGCT GGCCAGCGTG TTGCCTTTCC AAATCTACTA CCATTGCAAC CGCCACCACT GGGTATTCGG GGTGCTGCTT TGCAACGTGG TGACCGTGGC CTTTTACGCA AACATGTAGT TCAGCATNCT CAGCATGANC TGTATCAGCG TGGAGGCTTC CTTGGGGGTC TGTAACGCT	P2Y Purino-ceptor 8	P55085	Proteinase-activated receptor 2 precursor	397	62	172	111	38
190711 (160444)	AA883367	Dbest	400	TTTAAAGGTC AGCCTCTTTG TATAGAAGCG GTGATGGGCG ATATCAAGTA TCTGGTGACA CTGATGCAGA AGAGCATGAA AGCAGTGTGG AAACAGGACA AAACCCCCAG AAAGGCAATC ACTTTGCAAG TCAGAGTCCC ATAAGTCCAG GTAGAGCCAT TTTTGACAGA GTTGAACACA AATGGGAAAC AAATTGCAGA TCTGAGGATA TCTGAACAGC AAAGATCCAA CAGGAAGTAG TAAGGTGCTC TATGCAAGGT CTTATCTTTC ACTAGCAAAA TGGAGATCAG GAGGTTGCCC ACCACGCTGA CTCCTATTAT GAAACCCAAG GAAGTCAGTT TCAGAAAGGC TGTTAGAGGC GAGAGATTTT GCAAAATGTT GTCAGCTGCA TGGCTATAGT	GPR85 SEQ ID NO:36	Q9Y5N1	Histamine H3 receptor	445	41	134	95	34

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
191218 (160457)	AF003828	Dbest	503	TTTGATTTCA TCACTGCAGC GTGGCTGATT TTTTTATTCA TGGTTCTCTG TGGGTCCAGT CTGGCCCTGC TGGTCAGGAT CCTCTGTGGC TCCAGGGGTC TGCCACTGAC CAGGCTGTAC CTGACCATCC TGCTCACAGT GCTGGTGTCC CTCCTCTGCG GCCTGCCCTT TGGCATTCAG TGGTTCCTAA TATTATGGAT CTGGAAGGAT TCTGATGTCT TATTTTGTCA TATNCATCCA GTTTCAGTTG TCCTGTCATC TCTTAACAGC AGTGCCAACC CCATCATTTA CTTCTTCGTG GGCTCTTTTA GGAAGCAGTG GCGGNTGCAG CACCCGATCC TCAAGCTGGC TCTCCAGAGG GCTCTGCAGG ACATTGCTGA GGTGGATCAC AGTGAAGGAT GCTTCCGTCA GGGCACCCGG AGATTCAAAG AAGCATTCTG GTGTAGGGAT GGACCCCTCT ACTTCCATCA TATATATGTG GCTTTGAGAG GCAACTTTGC CCC	SEQ ID NO:39	P35410	MAS-related G protein-coupled receptor MRG	378	235	347	121	38
160458	AI264302	Dbest	491	TTTAAATATA AAAACTTTAT TGGAATAGCA TGTTAGCAGC AGTGAACAGG GCATGGCACA GAAGGTTTCC AAAACAAGTT TAGCATGAAG GATGCCATAT GCTGTTGCCA ACAACTAGAA CACGGTGACT AAAGACACAG TTCTGAATGT CCAGCACAAC CTCTGGCCTG CAACTATGTT CAGTGATGAT GATAAACAAG GTGGTGACTT GGAAGGAATC CCTATGTCAA GTGAGAAAAA AAAATGATGT CTGACCTCCT TATATATGTA AAAAATATAC CTTCAGAGTC CGTCAGTAAG CTGGAAGAAG TGGATGTTGA AGTTTTTAAC ATCGATGATG GGTCTCCAGT TGTTCATCAA CCCATGGTGA AATAGCTGAA CGGTTCTGAA TCAAAGGTGA TCCTAATAGT GAAGACATTA ACATTGCAGA AAAAGTGCCT ACAGATTATA TGGTGAAAAT ACGTGATGGG CTTCTTGAAG GACTAGAGCA G	SEQ ID NO:10	U45983	CCR8 chemokine receptor (CMKBR8)	1944 bp	1585	1690	108	88
191168 (161362)	AA274112	Dbest	542	GGCCTCTGAG AACCTTGGTG TGCCAAGTTA CTTCAGTCAC ATGTTATTTT ACAATGTATA TCAGTATATC GTTCCTGGGG TTGATAACCA TTGACCGCTA CCTGAAGACC ACCAGGCCAT TTAAAACGTC CAGCCCCAGC AATCTCTTGG GTGCAAAGAT TCTTTCTGTT GTCATCTGGG CCTTCATGTT CTTAATTTCA CTGCCTAACA TGATTCTCAC CAACAGGAGG CCAAAAGATA AGGACGTAAC AAAATGTTCT TTCTTAAAGT CAGAGTTTGG TCTAGTTTGG CACGAAATAG TCAATTACAT CTGCCAAGTC ATTTTCTGGA TTAATTTTTT AATTGTCATC GTTTGTTATA GCCTCATTAC CAAAGAACTC TATCGGTCTT ATGTCAGAAC AAGGGGTTCA GCCAAAGTTC CCAAGAGAAA GGTAAACGTC AAGGTTTTCA TCATCATTGC TGTATTCTTT ATTTGCTTTG TTCCCTTCCA CTTTGCACGG ATTCCCTACA CCCTGAGGCA GACTCGGGCC GTCTTTGACTGC	P2Y12 platelet ADP receptor SEQ ID NO:38	Q15391	ORF	338 bp	89	267	179	45

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
162615	AI479284	Dbest	508	TTGAAGCCAC TGAGACATTC TTGTTTTATT CCCAGACCCC TAAATCAGAA AACCCGATCG AATACTGAGC ATAATTTCTT CATTGACATT TGTCTCTAAA TGTCAAGTTG TTCTGGAAAT TTTTTCTTGA TTTTATGATT CCTTGCCTTA TTCATTTGAG ACAAACTGAG TTAGCATGAT GTTGTCGGAG GAATCTCCAG TATGAGAAAA TGCATAATGG CCTTTGTTTT GCAGTGGGTT GAAAGGCTTT GAGAATTTGG GTTTGGCAGA TAAATCTGAT GAGTTTTGCT TTTCTGTTTG CTTCCAAGAA CTTAAGGCAG ACAACTTGTT GAACAGAAGT TGTCGCAGCT TACTGTCCAA GAGTATTCCA AAGCATAAGA TAAAAAATCC CTGGAATGCA TTGAGTAAAG CAAAAATAAC ATGCCAAGCC AGATTCTGGC TGTCCACTAT TGTTCCTATT CCAAAGCCCC AGGTGAGCCC TAGCAGAGGG GTCAGAATGA GGAGGCTC	SEQ ID NO:12	O94858	KIAA0758 protein [Fragment].	986	860	961	104	47
168928	AA551068	Dbest	343	GGGGTCTTCT AATGTGACCC TGAGCAATCT TCTGCATACC AGTAAAGACT GTTCACTTTT CCACCATGAA CTCCATCATC AGAAGACTGT TTCTTACTCT GTTTCTTACT CCAGATATGT TTTTCTTATA GGAACAATGC TGCTTTCAAG TGCATACAGA GTGGTCCTTT TGTTCAGGCA CCAGAAGAAA TTCTGATACT TTCACAGCAC CAGCCTCTCC CCAAGACCTT CCCCAGAGAA AAGTGCCACT CAGACCATCC TGCTGCTAGT GAGTTTCTTT GTGGTCATCT ACTGGGTCGA TTTCATCATC TCATGCACCT CAACCTCGCT ATGGGCATAT GAC	Expressed in colon metastases SEQ ID NO:41	Y17566	M21 pheromone receptor, Mus musculus	959 bp	700	752	53	88
189873 LG155	AC007104	Genomic Clone	852	GCTCCTCAAG CCAACTGCCC AGGGAGCGCC CAAGTCCCTG GCCTGGGCCT GTCGGCATGC ACTGGGCCAA TACCTGGGGT GCAGGTCGGC GAGCAGCGCG AGCGCCTTCA TGGTGACGGT GTCCATGCGC TGGCAGTGTC TGCGTGCCAC CCGGTGCACC TGGAGCGAGG TGAGGCAGAG CACCGCCAGC GGCAGCACGA AGCCCACGGC ATGGAGCGTG GCGGTGAAGG CTGCGAAGCG CGGACGCTCA GGCTCGGGCG GCAGGCGCAG CGAACAGGAC GCGAAGGCGC TGCTGTAGCC AAGCCACGAG CAGCCAAGTG CAGCGCCTGA GAAGGCCAGC GACTGTCCCC AGGCACAGCC CAGCAGCAGG CCGGCATAGC GCGGTCGCAG GCGTCCGGCG TAGCGCAGTG GGAAGCCCAC TGCCAGCCAC TGGTCTGCGC TCAGCGCCGC CACGCTCAGC GCCGCGTTGG ACGCCAGGAA GGTGTCCAGG AAGCCAATGA CTTGGCATGC GCCGGGCGCC GACGGTGTCC GCCCGCGCAT CACACCGAGC AGCGTGAAGG GCATGTCCAG CGCCGCCAGC AGCAGGTGGC CCAGAGACAG ATTCACCAGG AGGACGCCTG AGGCTCGAGT GCGGAGCTCA GCGCTGTAGG CGCAACAAAG CAGCACCAGT GCGTTGGATA GCAGCGCCAC GGCCAGTACC ATCACCAGGA GACCCGCCAG CAGCGCCTCG CCGGGGCCCA TGGCGCTAGC GGCTCGCCAG GCACCCTGGG GTTCTCATGG CTCTGCTTCG GGCGCGAGCC TGGGAAAGTG AGGCGATGGA GCAGCTGAGC GGCGCGCCCA CGGCTTCTGG GA		P21918	D(1B) dopamine receptor	477	28	266	221	27

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Repre-sentative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
189874 LG349	AC008571	Genomic Clone	813	TCTAAGTTTT TCTCTGAACT TTGAGCCTGT GAAAAAAGAA GGGATGCTGC CTCAGGCCAC CCCAGCCTAG ATACTCACTC TGAGTGCCAT GAGGTAGTAG AGGACACTGA TGACAGTCAT GGGGAGGAGG TAGAATAGGA AGGAGGTGAC CTGGATGATG AAATTGTAGA TCCACATGGG CTTGATGACC GTACAGGTGG CCGAACCTGG GACCAGGGAC CCATTGGGGA AGTAGTGGAA CTTGATGCCA TGGATGCTGG TGTTGGGCAG GGAGAAGAGC ACGGAGAAGC CCCAGACGAT GCCGAGGATC CTGAGGGCCC GGCGCCGGGT GCTCTGCAGT TTGGCGCGGA ACGGGTGTAG GATGGCCACG TAGCGCTCCA CGCTGACGGT GGTGATGCTG AGGATGGAGG CGAAGCACAC GGTCTCAAAG AGGGCCGTCT TGAAGTAGCA GCCCACGGGC CCGAACAAGA AAGGGTAGTT GCGCCACATC TCATAGACCT CCAGGGGCAT TCCAAGGAGC AGGACCAGGA GGTCAGAGAC CGCCAGGCTG AAGAGGTAGT AGTTGGTGGG CGTCTTCATA GCCTGGTGCT GCAGAATCAC CAGGCACACC AGGACATTGC CAATGACCCC CACCACAAAA ATTGGCACAT ACACCACAGA CACGGGGAGG AAGAAGTGGC TGCGCCGAGG TCCGCAGAGG AAGGCCAGAT ACTCCTCGGT GCTGTTCAGG TGTTTCTGGA ATGGATCTTC TAGTTTCTGC TGGTAGATCC AGGAAGCATT CTGAAGTTTT TCCATCCCTG ACATTAAAAT CCA	Neuo-medin U receptor 2 SEQ ID NO:14	O43664	Orphan G protein-coupled receptor	403	17	235	218	56

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
189876 LG1543	AP000808	Genomic Clone	1113	GGGCATGGGC TCGCACAGGT GGGAAGCACC TGTGGGCGGC TCTCAAGCCC CCATCTCATT GGTGCCCACG GTGGGCGTCT CCCCACCTTC CAGCTCGGGC TCCTCGCGAA GCGCCTGTTG GAGCACAGTC CCCAGGGACC TGGTGGGCAG CCTGTGGCTC CTCCGGCTGC CCACCAGGAA GTAGATGACG GGGTTGGCGC TGCTGCTTAC GGACGAGGAG AGGCGTGACA AGCTGAAGCA CAGGACCTGC ATCTCGGGCG GCAGGCTCAA CCAGTAGAGC ACAAACCAGT AGATGCTCAG AGGCAGGGAA CAGATGAGGA ACACCAGGAC AGAGGCCAGG ACCACCACGA ACAGCCGTGT GGGCTGCCGC CGCCACTGCT GGGAGCTCCT CCGCACCCAG ACAAAGAGGG TCAGGCTGGA CAGAGTCATC ACTGGGGTTA AGACCCCCAT GATGAGGGCG GCCTGGACCA TGTCCACCCT GAAGCACCGA TCTTCATTGA ATTTCAAGAA CTTGCTGCAG AAGGAAGAGG TCAACCCGTT CATCAGGAGA CAGAGTGTCC ACAGCAGGCC ACACACCCAG GCTGACAGGT GCCTGGGCCG GTGACACTTG AACCAGATAG GGAAGAGGAC AGAGAGACAG CGCTGGGTGC TGATGGCCGT CAGCAGGCTC AGGCCCACTG TGTAGGCAAA GTACATCAGT CTCTTCATCA GCTCGTGGAC CTTGTCAGTG GTATTGACCA GGGGCTGGGT TTCCAGGCTG AGCGTGGAAG CCATGCTGAA GAGGAAGAGG AGGTCGGCTG CCGCCAGGTT GAGGATATAG ATGCAGAAGG GGTTCCTGTG CATTCGAAAG CCCAGCAGCC AGATCACCAT GCTGTTGCCT GCCATCCCGC ACAGGCAGGT GAACATGGCC AGGGAGCTCA GCACCAGGTA GGCCGTGTGC ACTGTGCTCC CTCTGGAATA GTTTAGGGCT GACTCCACGG TCCCACTGCT ATTCAAAGTC TGGTTCATCC CTACGAGAGG AAGATGTACC AATGTGAAAT TCTGTGTTGC TGGGACCACG GGGGACCCCT GGGTGCCCCT CGAATTTCCA GCTTCAGAGC TCTCCCCTCC AGG	SEQ ID NO:15	P04201	MAS proto-oncogene	325	7	306	297	35
189878 LG1143	AC016362	Genomic Clone	504	CCTGGCAGTG CCGATGTTCC GATACTGGCA CAGCAGCAGG TGCCGGAAGG TCTTTTTAAA GGTGGCGTTG CACAGAGCAT AGCAGGCAGG GTTGATGGTG CTGTTGACGT AGCAGAGCCA GTAGCCAATG GACCACACCG GGTCAGGGAT GCAGCTCTGG CAGAAGGTGT TCACCAGGAC CATGACGTTG TGAGGCGTCC CGGTGAGGAT GAAAGCTAAC ANAATGGCAA AGATCGGTCG TGGCACTTTG CGCTCCCGGG CCCGCATCTG CCGCTTCTTG CGCACCTGGG TGCGAGCGAT GCTAGCGAAC TTGCGGGCCA CGTTGGCCGC AGGCGCATGC CAGNCGGCGT GGGAGGGACA ATCTCAGGGC TGGCACACAC TCATGGGCTG GCTGGCTTCG TCAAATTTTG GATCTTGGAC CATCTGGGAG GCTTGGTTGA AGGCCCCCGG CTCGGACTTG CGGGCATGAA TCCAGGCCTT ACTCTANAGG ATCCCCCCCT CTCC		P08173	Muscarinic acetylcholine receptor M4	479	369	479	111	92

LS Cluster ID: Current (Original) LG NO.	Acc. No.	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
189884 LG608	AC011375	Genomic Clone	1137	AATCTCTGCC TTCTCAGTTT TCCCTTTGCC AGAGGAGGGA GAGCTGGGTT TCTCTTTTTC TGGTATGGAT GCTGGGGATT CTGGAGATGG AACCTTGTCA GGAAGACCCT CTGAGTTGCC AGCTGGTGTT TCCTGAGACT CTGAGACAGT TGGAGGTTTT TTGGTTATCA TCCATTTCCA TACACCTTTC AAGCCTTCCC TGAACTCTTC CGACATCACA AGAAAAATGA GAGGATTTGC TGAAGAGATG GAAAACATCA AGACTTGAGA CAGGGCTATG AAACCTTGTG GTGGGGCCGG GCCTGCAGCC TTCAGATGCC ATACCCACAG CCAAGCTACC CATTCGGGGA GCCACAAGAG AGCAGAGATG ATGGCAATGC TCAGCAGCAT CACTGTGACT TGCTTTGAGC GTATCTGGTT TCTAAGATTT TGAGTCTTAG TTCCTCGTTT TTTACATTGG TCATAAGCTC TCCAGAAATA AAAGCTGGCA AAAAATAATG GAAGGCCAAA TGCCAGGAGT GGGTAGAGCT TACCAAACAT CGACATAAAC TCTTCAGCCA CAGCTGGTAC ATCCACGAGG CACATTTCCA CACCTTCATG ATGCCTGATG GTGCTAAAGA ACCATTCCGG CAGGGGTAAC AGGCTAGCCA CAGTCCAGAT GGCCACCAGC ACTGACCAGA TGGTGTAGTT GTGGATACTC ACTTGCTTGG CTGGGTCACT TGCATACATG AAGCATACTT TGGCCACCAC AACGATTGTC AGGCTCTTGG CTGCCATGCA TGTGTGGATA AACCAGTCAG AGGACTTGCA GACAAACCAG CCTAGATCCC AAACACTTTT GGAGTACGCC GTAGCTCGGA TAGGTGCAGA AAACAGCAGG AGGGAGAGAT CAGCCAGGCT GAGATTCAGA ATCAGGGAGT GGATCATGGA TGGCTTTCCT TTCCAAGCAT TGTGAAGGAG GATGCCAATC ACACACAGGT TTCCCACGAA GCCCACCAGG CAGACAGCCA CCAAGAGAGC CGGGATGATG GTTCTCCAGT CCTGGGAATC AGAGGGCAGG TACCCTCCGG CAAAGTGGAG GTGAGCAAAG GACACATTCA TGCTGCTGGA GTTAGAGTCT GCAAAGGCAG CTGCCAG	SEQ ID NO:40 Amino acid sequence: SEQ ID NO:18	Q9NS07	GPR SALPR	469	81	444	326	28

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
189879 LG1390	AL049739	Genomic Clone	330	CTCCATCTCA CTCAACAAAG CCCTCTGAAA AACAGCCTTT AGGGACTCCC TGTGCCTCTT CCTGTAGAGT TACTCAGCCA AGAAGTAGAT GACTAGGTGA GGCATGCTGA CCACAATGGA CAGTAGCAAC AGGAGGTCAA AGGCAAGGGT CAGAAACTTT CCTGGCAGGC ACACAAGGAC AACTAAGGGC AGGACCCAAA GGAAGAAGCT GATGATCACA AAGCAGACAA CATGATAGGT CCTGATGGGG GAACACCACT GGGGACAGCA CAGACCCCAG ATGATCAGGA TCAGCTTGGA CATGCCCATT ACAAAGCAAA TAAGTACATG		P35410	MAS-related G protein-coupled receptor MRG	378	232	342	110	39
189881 LG1391	AL049739	Genomic Clone	492	ACATGCCAAG CACTTCTCCC AGTACTCACA GACCTGGCTA ACTGCATACA AAGAAAGGGC CAGGGCCCAC CTCACCATGG CAGAGGTGTG CTCTGGGCGG TGGCAGCACC AGGTGGGACA GAGGGCACAG AGAAAGCTCT CAATACTCAT GGCCACCAGG AGACAGAGAC CCACTGTGTC GGAGAAATAG GAGACAGGAT CCAGAAACAC AGCCACCTGC AATGCCGCCT GGTGATACAG CATGAGGATT TTCTCCAGCA GGATCACAGT TACACAGGAG AGGTTGACCA TATCAACAGT GGCCAGGTTA AGGATGTAGG TCACATAGGG GCTGCTCCAG ACCTGTGAGT AGAGAAGCCA GCAGATCACA TCATTGCCTA CCAGTCCACA GAGGGCCACC AGCACTGTCA GGGAGAAGAC CACCTGCCTG TCCACCAACC ACTCACCTCC CGTATGGCTC ATGTTCACAT GTCCTGAGGT CTCAGTCTCA TT	SEQ ID NO:16	P35410	MAS-related G protein-coupled receptor MRG	378	54	221	164	39
189884 (189882) LG610	AC011386	Genomic Clone	429	GGAGGGTACC TGCCCTCTGA TTCCCAGGAC TGGAGAACCA TCATCCCGGC TCTCTTGGTG GCTGTCTGCC TGGTGGGCTT CGTGGGAAAC CTGTGTGTGA TTGGCATCCT CCTTCACAAT GCTTGGAAAG GAAAGCCATC CATGATCCAC TCCCTGATTC TGAATCTCAG CCTGGCTGAT CTCTCCCTCC TGCTGTTTTC TGCACCTATC CGAGCTACGG CGTACTCCAA AAGTGTTTGG GATCTAGGCT GGTTTGTCTG CAAGTCCTCT GACTGGTTTA TCCACACATG CATGGCAGCC AAGAGCCTGA CAATCGTTGT GGTGGCCAAA GTATGCTTCA TGTATGCAAG TGGCCCAACC CAGCAAGTGG TTTTTCAACT ACCCCATTTG GTAATGGCGG TTGGCCTTTT GACTGGGGCT TACCTGTTA	Putative GALR4 receptor SEQ ID NO:1	P47211	Galanin receptor type 1	349	46	130	85	40
189883 LG455	AC009763	Genomic Clone	432	CAGGTCAGCC CCTCCTCCAC ACTGGCCAGC CTCAGCGACA GGACCCAGTT CCCATCCTTC TTTCAGACCC TGCTCAGTCA CCTCACGACC ACCCATGCAG TGGTGCAGCT GATGCTTCAC TTCCGATGGT CTTGGGTGAG CGTCCTGGCG CAGGGGGACG ACTTTGAGCT GCAGGGCAGG TCTCTGGTCG TCCAGGAGCT GGGCCAGGCT GGGGTCTGCA TTGAATTCCA ACTTTGCATC CCCACCCGGG AGTCCCTGAA GATGAAAAAC ATCATCTGGC TGATGGAGAA CTGTACGGCC ACCATCATCC TGGTTTTCCT TAACAGCTAG CACTTCAGAC TCATCCTGCA GTGCCTGGCA GGCCGAGGCA TCCTGGGCCA GGTCTGTGTC AGCTGGGACA CTCTGCACAT GGCGCTGGTC CTGACTGTGC CA	SEQ ID NO:17	P41180	Extracellular calcium-sensing receptor precursor	1078	164	308	144	25

LS Cluster ID: Current (Original) LG NO.	Acc. No.	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
189885 LG5574	AC011402	Genomic Clone	963	TCTGCTCTTG ACATCTTTTC CATGATCATC TACACAGTGA CTTTCTTCCT AGGCTTGGCT GGCAATGGCC TTGTCATTTG GGTAGTTGGA TTCCACATGT CCTGCACAGT CAACACGGTG TGGGTGGTAC CTCAACCTGC CCGTGGCTGA CTTCATCATC ATCTTCCCAC TGCTTCTCCA GCTGGTTATG GTAGCTCTGT AACCCTTTGG CCAGCTGCTC TGTAAACTCA ATAGCACCAT GTCTATTTTT AACTTTCTGG CCAGTGTCTT CCTTCTGACC CTCATCTCCA TGGACCACTG ACTTGTGATC CTGTGGCCAA TCTAGTCCTG GAACAATTGC ACACCAGCAA AGGCAACTCT GGGGCCCTTG AGGACCTGGC TTTTGGCAAT TTGTTTCTCT GTTCCCTACT TGATCTTCAA GGAAACTCGT GGTGGAAAGT GTCACCCTCT TTGTACAACC AGTATGATCT GCAGAATGAA ACTCAAGGAA GTCACCAACT TTGGAAAGAG ATTATCATTC CATGGCACCA AACGCTGGTC ACAACAGCCC ACTTTTTCTT TGGCTTCTTT CTCCCTCTGG CTATCATCAC TGGCTACTAC ATCCTTGTAG CCTTGAAGTT AAGAGAAAGG CAGCTGGTTA AGTTTAGCTG ACCCTTCCAG GTCCTTGCGA CTGTGGTAAC CACCTTCTTC CTCTGCTGGT TGCCCTTGCA AGTGTCCCTG TGGCTGGACT TCACATCATT TCGGGAAGAC AGAGAGGGCC TGAACCAGGT GGCCTTACTC CTAAGACCCC TGGCCTTGTC TATGGCCTTT ATCAACAGCT GTCTCAATCC AGTTCTCTAT GTCTTCATTG GGCATGATTT CTGGGAGCAC TTGCTCCACT CCCTGCTAGC TGCCTTAGAA CGGGCACTTA GTGAGGAGCC AGATAGTGCC TGAATCCCAG CTCCCAGGCA GATGAGTCCT TTA	SEQ ID NO:19	P21462	FMET-LEU-PHE receptor	350	26	62	37	56
189886 LG1121	AC016189	Genomic Clone	330	GGGGTCTACC TCATGGCCTG TGTGAGCGTG GACCATTACC CAGCTGTGGT CTGTGCCCAC TGGGGCCCGC GCCTCCGCAC GGCTGGCCGC GCCAGGCTGG TCTGCGTGGC CATCTGGACC TTGGTGCTGC TGCAGACGAT GCCCTTGCTC TTGATGCCCA TGACCAAGCC GCTGGTGGGC AAGCTGGCCT GCATGGAGTA CAGCAGCATG GAGTCAGTCC TCGGGCTGCC CCTCATGGTC CTGGTGGCCT TTGCCATTGG CTTCTGTGGG CCAGTGGGGA TCATCCTGTC CTGCTATATG AAGATCACCT GGAAGCTGGG CAGCACAGCT	SEQ ID NO:20	P32249	EBV-induced G protein-coupled receptor 2	361	118	227	110	40

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Repre-sentative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
189887 LG626	AC011457	Genomic Clone	537	GATGCATCCA TGTAACCCAG TACTGGAGGT CAAGATGGAG ACAATCTCCA CAGCCACCAT GGCCTTTCCC TGAGTGCTAC TATAACTCAG GAGGAAAGCC ACCCAGATAC TGCAGAACAC CAACATGCTG AAGGTGAGAC ATTTGGTTTC ATTGAAGATG TCAGGCGGGT TACTTTCCAG GAAGACTACA GCAAAGGTCC CCAGAACCAA GAAGCCCAAA TAGCTCAAGA CACAGTAGAA AACAGTGATA GAGCCCCCAT TACACTGGAT GATGATTATA CCAGGACTAT GGGTCTTTGT CTAGGGAGGA AGGAGATGTT TCCAACTAGA TTCCAGAGAT GGTCACTTGG ATGAGGGGGC AAAGGTGACA ACAGAATGAG ATTTTAGGGT ACCCAACAGG ACCTTGGGTC CTGGCTGGCC TTGTAGCTTT CCCATCCTGC TAGCTGGCCT TGTAGCTTGG AAGGCCATGA CCACAGTGAA AGTTTTGGCT GAAGCAGAAG ACACAGCCAC TGTGAACTGA ACTGAAAAGG TGGCCTG		AAD14370	Calcium receptor|CaR protein (fragment)	266	39	210	179	31
189888 LG5533	AC010896	Genomic Clone	1317	CCAGCCCCAG GTGAGGCCAA AGATGGGTGT AAGAATGAGC AGGGCTTTGA TCACCCCCAG CAGAGCTTGG CGCTTCTCTG CTGGGGGTCC CTCTGACAGC GAAGGTCTCA GCAACTTCAG CATGGCCATG GCTAGTACCA GCCCATTCAC GCCTATGATG GCCAGCACTG GCCCACGAAG GTGTATAACG CCCCTCCCTT CCCATCCAAC CAGCATTCCC CCTCCCTCAG GTATTGCCCT TGAGGTAGGT AGAGCCCCAG GGTGACACCT GCCAACCCCA GTGGGCACAG GTAGCCCAGG AGCACCATGA GGGGGAGAAC TCGGTGCTTT GCCAGCTGGT GAAAGACAAA GAGCAGCTGG TGGGCCAACA CCAGGGCCTG CGCCACATCC AGAAAAAGGT GGCCAGGTAG AGGAAATGAC AGAGGAAGGC GGCACAAGGC AGAGCGGGCT TCGGGGCCCT GGAGAGAGGA ATGGGGCGCC CAGGAAGCAA GTGTCTGCGG CCAGCAAGCA GAACACCATG TTGAGCAGGG CGGCGTGGCG GAAATAGGAG ATCTTGTTCC GCACCACGAC TCTCCACACC AGCCAGTACA CACCCAGGCA CACAAGCAGC GCCAGTATGG AAGCTCCCAA GCCCACTTGA GTCAGCAGCG CCAGAGCGGG TTCTTCCGGA ACAGTGTGTG GGGACATGAG GACGGAGAAG GCAGTGAGGT GCTGGCAGAG GCACTGAGCA GTGGGGCTGG CACTGGCCAC CTGTGCCTGG CACCCTTCTT TGGACCAACC CCCCCTGCCC TGGAAGAGAC TGTGATCCCA GAAGACACAG TGAGGGGAAC CATCTGTGTT CCCAAAGTCC ATGATGACCT CTCCCTGGCT GAAGGCCCGG TCACCTGCCA TGATGGAAAT GACAAGGACC AGGCCAGGAG TGGCATAGAG GGAATCCCCC AGCCCTTGTC CATAGTTTGA GGGCAGAAGG TGGTCCAGTT TTCGCAGCAC CAGGCTAGTA ATACTTATTT CAGTTCCATT ACGGACCAAT GGGGCCAGTG AGTGCCTGGG AATCTGAGCC TGCAGTGGGG GCCGAGTAGG GAAGGAGATG CTGTAGTCAG CAGGAAACGT GGGTCCAAAC AGCTGGCTCT GCAGCAGCAC ATTGGGTAAG CTGAAGGCGA AGGGGTGGTC CTGTGGGCAC AGGCTGCATG CCAGGGTCTC CACAGCCAGC AGGAGAGTCG AGCCTGCCCA GGGCTTCCGG GCTTGGGCCA GGGTCCACAG AGACCTGGTG TCCATATCTA GGACCTTGTC TGTGGCAATC AGGAGATTCT GCAGCAC		O94858	KIAA0758 protein [Fragment].	986	431	874	439	26

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
189889 LG1183	AC016856	Genomic Clone	420	CATTTGGTCT TTTGTGTTCT TTTCAGGATG GAGAAGGATG ATGTAGCATT TGGAAAAGAA GAGGGAGGAC ATGAGGCCTC CTGCTGATGC CAAGATGACA AAGACCTCCA TGGCCATGGT GTCTTTGCTG TGGGCATGCA TGTGGGCAGG TATGAAGGAC ACCCAGACAC AGGAGCAGAC AACATGCTGA GAGTGATATG CTTTGCTTCA TTGAAGGTGT CAGGCAGCCG GCAGGTGGGG AAGGTCACCA GCAAGCTGAC CAGATCTAGC AAACTCAAGT AGCCCAGTTC CAAGAGAAGT TCCAAGGAGC CTTTATCACA CTTTACAGTC ACTGTGGGCC ANGGTTCTGT AGAGTTTCAC AGGCTGAGGG GGCCATCTGG TCACCCAGAG TGCACACAAG GCTGCCTGGG TCAGGGAACA	SEQ ID NO:21	AAD14370	Calcium receptor|CaR protein (fragment)	266	89	230	140	30
189890 LG1182	AC016856	Genomic Clone	612	CCAAGTTACA AGGCTGGTGA GAACATTTGT GGATCCTGAG CAAGTGTGTT CAGGCCAGGA ATGTACAGCT CTGTTCCTTT CCCCCAGATC AGCCATGGAT CTCAGCGCCC CCAGTTTTCC AACTGTTTCC AGTTCCCATC CTTCCTCCGC ACAGTGCCCA GCAGCACCCA CCAGCCTTGA GTCCTGGCCA AGCTCTTGAG TTACTTTAAC TGGACCTGAG TAGGCCTGGT CAGTTATGAC AACAGCAACT TTGAGTGGCT GGATCAGCAG CTGCAGAAGC AGATCGGGGG CGAGGGTGGC TGCATGGCCT TCTCAAAGAT CAGCAGTGTA GATGACAGCA TCAACAGCAT GGCCACCGTC ATTGCCTAAA CCCCTGCTGC CACCGTCATT GTCTGTGACT GCTACCATTT TCACTTCAGA CTCCTGGCGG GGGCCCTTCA GGAGAACAAT GTGAATGGGA GGATGTGGAT CTTTTCCACC TCCTTCACAT ATAACCCCTC GGTGCTGGGT CCCAAAGCCC ACGAGTTGCT GAATGGCAGC TTGAGCCTGA CCATACACTT TAGAATAATA CCTGCCTTTA AGGACTTCCT GTTGGCGCCT GCGCCCAACC GT	SEQ ID NO:42	P41180	Extracellular calcium-sensing receptor precursor	1078	136	341	204	25
160833 (189891) LG5616	AC011638	Genomic Clone	419	GGAGTTGTTG TCTTTCTGGT TTGTTCAGCT TCTTCCTTGT TTTTCTGTTT TATTTTTTAA TTCCATTATT TTCTTTGTTC CAAGGTAAAG GAAGCACCTG ATGAGGCTGA AGTGGTGTTC ACGGCACATT GCAAAGCAAC GGTGTCCTTC CAGGTGAGCG GGGACTCCTG GCGGGAGCAG AAGGAGAGGC GAGCAGCCAT GATGGTGGGA ATTCTGATTG GCGTGTTTGT GCTGTGCTGG ATCCCCTTCT TCCTGACGGA ACTCATCAGC CCACTCTGTG CCTGCAGCCT GCCCCCCATC TGGAAAAACA TATTTCTGTG GCTTGGCTAC TCCAATTCTT TCTTCAACCC CCTGATTTAC ACAGCTTGTT AACAAGAACT ACAACAATGC CTTCAAGAGC CTCTTTACTA AGCAGAGAT	SEQ ID NO:11	P47898	5-hydroxy-tryptamine 5A receptor	357	247	340	95	74

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc No.	Homolog Name	Len.	From	To	Aligned	Percent
189885 (189892) LG606	AC011352	Genomic Clone	954	GACATCTTTT CCATGATCAT CTACACAGTG ACTTTCTTCC TAGGCTTGGC TGGCAATGGC CTTGTCATTT GGGTAGTTGG ATTCCACATG TCCTGCACAG TCAACACGGT GTGGGTGGTA CCTCAACCTG CCCGTGGCTG ACTTCATCAT CATCTTCCCA CTGCTTCTCC AGCTGGTTAT GGTAGCTCTG TAACCCTTTG GCCAGCTGCT CTGTAAACTC AATAGCACCA TGTCTATTTT TAACTTTCTG GCCAGTGTCT TCCTTCTGAC CCTCATCTCC ATGGACCACT GACTTGTGAT CCTGTGGCCA ATCTAGTCCT GGAACAATTG CACACCAGCA AAGGCAACTC TGGGGCCCTT GAGGACCTGG CTTTTGGCAA TTTGTTTCTC TGTTCCCTAC TTGATCTTCA AGGAAACTCG TGGTGGAAAG TGTCACCCTC TTTGTACAAC CAGTATGATC TGCAGAATGA AACTCAAGGA AGTCACCAAC TTTGGAAAGA GATTATCATT CCATGGCACC AAACGCTGGT CACAACAGCC CACTTTTTCT TTGGCTTCTT TCTCCCTCTG GCTATCATCA CTGGCTACTA CATCCTTGTA GCCTTGAAGT TAAGAGAAAG GCAGCTGGTT AAGTTTAGCT GACCCTTCCA GGTCCTTGC ACTGTGGTAA CCACCTTCTT CCTCTGCTGG TTGCCCTTGC AAGTGTCCCT GTGGCTGGAC TTCACATCAT TTCGGGAAGA CAGAGAGGGC CTGAACCAGG TGGCCTTACT CCTAAGACCC CTGGCCTTGT CTATGGCCTT TATCAACAGC TGTCTCAATC CAGTTCTCTA TGTCTTCATT GGGCATGATT TCTGGGAGCA CTTGCTCCAC TCCCTGCTAG CTGCCTTAGA ACGGGCACTT AGTGAGGAGC CAGATAGTGC CTGAATCCCA GCTCCCAGGC AGATGAGTCC TTTA	SEQ ID NO:19	P21462	FMET-LEU-PHE receptor	350	26	62	37	56
189893 LG699	AC011647	Genomic Clone	720	AATGGCCACT TTGGGATGTG GTCCTCCTGG AGGTAAGGAA GAGGGCACAC TTCTACCAAT GCTCTGCCTA GCCCTGGGGT CTCAGGGTCT GGAAGCCTCG TTAGGTTTGA CAGGTTTTGG TGGAGCATGA GCCCCAGAGG GCCTCCTGGT GGGTGGTCTT GGTAGCCTGG TCTTCATTGT CAGGAAGGTG AGGTCATAAT TTATCAGGTA ACCATCACTG AAGACATACG GCTTGTGGTC TGAGGGGCAG TAGTTGATGC CCTGCAGCTT TTCCAGCATC TTGTCCAACA GGATGCTGAG GCGGCGCTCC TGCCCGGTGG TGGTGTCAAA AGCATAGAAG ATCTCCTCTT GGTGGGTGTT CAGTGAGTGT AAGGCATAGA GCACCCCACA GGCCATGAAG GCCCCTGACA GGGCTGGCTT GTACTGGCTG GTACGCCAGG TTTTCTCCAC TTCTAGGGTG CTAGCGTTGA GACGACTCAC AACCAGGTTG CCCTTGCTCT CCTCAGTGGC ATACAGAACC CACAGCCCCT TCTCATCACC AGCAAAATCT AAGTCCTTCC AGGGCACACC AGCACAGGAA AAGCGGTTGT TATAGGTGGC ACCAGGCAGC AGACGCCACA GCACCAGTGT GTTGGAGGAA AGGTCCATTT TGGCCATGTC ACTTGTGCCA CAGTAGTTAA AGTACATAAA GTTCTTGTAC ACAACGTTTC CACTGCCATC	SEQ ID NO:43	O94910	KIAA0821 protein	1474	212	447	240	33

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
190701 (189894) LG1446	AL121834	Genomic Clone	660	GCTTGCATAA GCATAGACAG ATATGTGGCA GTAACTAATG TCCCCAGCCA ATCAGGAGTG GGAAAACCAT GCTGGATCAT CTGTTTCTGT GTCTGGATGG CTGCCATCTT GCTGAGCATA CCCCAGCTGG TTTTTTATAC AGTAAATGAC AATGCTAGGT GCATTCCCAT TTTCCCCCGC TACCTAGGAA CATCAATGAA AGCATTGATT CAAATGCTAG AGATCTGCAT TGGATTTGTA GTACCCTTTC TTATTATGGG GGTGTGCTAC TTTATCACGG CAAGGACACT CATGAAGATG CCAAACATTA AAATATCTCG ACCCCTAAAA GTTCTGCTCA CAGTCGTTAT AGTTTTCATT GTCACTCAAC TGCCTTATAA CATTGTCAAG TTCTGCCGAG CCATAGACAT CATCTACTCC CTGATCACCA GCTGCAACAT GAGCAAACGC ATGGACATCG CGATCCAAGT CACAGAAAGC ATTGCACTCT TTCACAGCTG CCTCAACCCA ATCCTTTATG TTTTTATGGG AGCATCTTTC AAAAACTACG TTATGAAAGT GGCCAAGAAA TATGGGTCCT GGAGAAGACA GAGACAAAGT GTGGAGGAGT TTCCTTTTGA TTCTGAGGGT CCTACAGAGC CAACCAGTAC TTTTAGCATT	C-C chemokine receptor 11 SEQ ID NO:34	P49238	Probable G protein-coupled receptor GPR13	355	123	304	182	38

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Repre-sentative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
189895 LG745	AC011780	Genomic Clone	1245	TGGCTCCTGC TTGGGGCTGG GTAGGGGTCG GGAAACCCAG GACTCAGAAG GACAGCCAGT CCCCTGAAGG AACTGCAGGA AGTTCTCCTC AATGGAGCCC TCCCGGCTAG GCAGCCTCAG CTCCTCCTCT GGAGCTGGCT TGAAGAAGCA GACAAACTGC TTGCTGAGCT CCCCCCGGAT CTGCCGGTTG AGACATCCAT AGAAGAAAGG GTTGGAAGTG AAGCAAAAGT AGCCAATCCA GGTGACCACA CTCTCCACCT GCCCAGTTGA AATGGGCTGA GCACTCAGGG CAACATAGAG GTGGAAAGAG AAGTAGGGCA ACCAACAGAG CAGGAACTGT CCCCCCACAG CCAGGAGAAC CACTGCTGCT TTCCCTCCCC CAAACGTCCG GTGTGGGGTG GTCTGGGGGG CCCCCGAGCT GGTGACCATC GTGGAGCGGC TGCTGAGAGA TTCGGAGCGT TGCCGGGGTG TCTCCATCCA CGTGGGCAGC GGCCCGTGCT GCATGGCAGC CACGCGGGCC ACTCGGAACA TGCTGCAGTA GACCACAAGT ATGAGGAGCA GGGGCAACAG AAAGTAAAGG ACAGCAAAGA CCACCACAAA AAGCTGGCAG TAGGCACTGT GGCTCCACTG GAGTGAACAG CCTGGGGGGA CACTGGGAGC TCCTTCCTCC CAGGAGACCC TTCCCAACAC TGGCACAGAA GCCATGGCCA AGGCCTTCAC CCACACACCC ACCAGCACAG AGGCCACCAG CCCCAGCGTC ATGCGCACCT CGTAGCGCAT GGGGTGGACT ACGTAATAGT AGCGCTCCAC ATTGATGGCT GACACCGAGA GGATGGCCAG GCTGACAAAG CACACGCTCA GAAACAAGTA GAGGCGGCAG GCCACCTCCC CAAAGAGGGC GTGGTCAAAG AGGGCAGAGC TGGAGAGCAT GGCCAGGGGC ATGAGGGTCA GGGCAGCCAG CAGGTCCACC AGGCAGAGGT GGAAGACGAA GACAAATTTT CGGAGGGCAG GCGTCTTGGC GATCACGGCC ATCACAGCGG CATTGCCAGC CACAGCAGTC AAGTCCAGCA GGAGCATGAA GAAGAGGGCC ACAGATTCCG AAGCAACATC CCGTAGCCCC ACCTCCGGGA CCCCACTGGC AGTAGAGGGA CCTGGGGTTT GAGGGACCCT CCCCAAAGTG GAAGAGTTCC CTGATGACTG GGGGATGGGT GAGGACTCCA TGGGGCCGAA AGAGGGCACC CAGGG	GPR61 SEQ ID NO: 22	P50406	5-hydroxy-tryptamine 6 receptor	440	2	337	340	32
189897 LG1440	AL121755	Genomic Clone	477	CACCCACCAT CCTCACCTGT CAATGGCAAT GGCCAGCAAG GCATTGGTGG AGACGTAGAG GGAGACGGTG CGCAGGTAGT TGACGGAGGC ACAGAGCACG TGGCCATGCT CCCAGGAGAG CTGCCGTACC ACGTAGTAGT CCATCTCGAA GGGGCAGCAG ATGATGGCCA CCAGGAAGTC GGAGATGGCC AGGTTGGCAA TGAGCAGATT GGTGAGGTTG CGCAACTTCT TATAGCGGGT GAGGGCAGCG ATAAAGACAA AGTTACCGAT GCCGCAGACC AGCATGATGC CTGCCAGTGC AATGCCAATG ACGATCTTGG CTGCGAAGAA GGTCCGGGTC TTGGTCATGT CCTCATCCTC ATCCATAGGG AGGTCATAAT CACCATAACT GAAGTTAAAG GAGAGGGAGG AGGCATGGTC TTGGGGTGGA TTAAAGTTGG GTGTGAAACT GGTGTTTCCA TTCTGGGCTG CCATGGT	GPR73 SEQ ID NO:23	AAD00248	Neuropeptide Y receptor type 2	381	31	148	120	34

LS Cluster ID: Current (Original) LG NO:	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
189897 (189898) LG1439	AL121755	Genomic Clone	792	TTGTGTGACA CCAGTGGGTC ACTTCAGCCT GATACAGTCC ACCTCTTCTG TGGTGGGCAC CCCGTTGGTT CTGAGGTCAA GGTCAGCACT GGACTTGCTC CCCCGCTGGG AGGGACGCCA GTGCAGCAGC ATCATCTTCT TGAAGTACTT CATGGTGTTG TTCTTGACCG TCACGAAGCA CACGGTGTTG ATCATGCTGT TGCTCATGGC GATGCACTCG ACCACGTAGA AGGCAGTGAG GTAGTGCTTT TCCTTCACGA ACACAGTGGG GAAGAAGTCA CGAACGATGG TGAAACCGTA GAAGGGTGCC CAGCACAGCA CATAGGCCGT GAGAATGCAC ATGAGCACCA GGACCGTCTT CCTGCGGCAG CGCAGCCGCT TGCGAATCTG CTCCGTCTGG AACCCAGGGA CTGCCTTGAA CCAGAGCTCC CGGGAGATCC TGGCATAGCA CAGGGTCATG GTGACCACAG GGCCCACGAA CTCGACACCA AAGATGAAGA GGAAGTAGGA CTTGTAGTAG AGCTGCTGAT CCACAGGCCA GATCTGGCCA CAGAAGATCT TCTCCTGGCT CTTGACAATA AAGAGGACCG TTTCTGTTGC AAAGTAAGCC GATGGATGG CAATGAGAAT GGACACCATC CAGACCAAGG CGATCAGGAA GGAGGCCGTT TGATAATTCA TCCGTGGTTT CAAGGGGTGA ACGATGGCGA GATATCTGGT GGGGGAGGGA AGCCAACAGT AGTAATGATG AATACGTGGA AACGTTAGTT TGTAACTAAC CCCAAACATA CA	GPR73 SEQ ID NO:23	P25103	Neuropeptide Y receptor type 2	407	130	317	193	23
3098 (189899) LG762	AC011923	Genomic Clone	303	GAAGGCCTCA ACCCCACTTT CGCCATAGTT GCCCTCGGAG GCCAGCGTGG ACACATAGTT CCATCCCAGT GCCCTCACGA TGTCCACCAT GGCCTGCGCC TGGTAGGAGT CGGGTGGCAC CACCCGGGAG AAGAAGTCAT AGCGTGTGGA GTCGCTGAGC TCCGGGGCTG TGGAGGCATA GCTGATCTGG GGTATCTGTG GGGCAGGAAG GACAGCTGGG CTGTGGATGG AGGTCAGTAA CTCAAGAGAG GGAGGGTAAG GGGGGCCCAG GACACGGACG GGGCACAGAA GGTGTGTGGC ATG	Metabo-tropic glutamate receptor 6 (SEQ ID NO:2)	O15303	Metabotropic glutamate receptor 6 precursor	877	157	233	77	89

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
189900 *LG629*	AC011461	Genomic Clone	1335	GGGGTGATGG TCAGTGGGTA GGGCCCCTTT CCCACGACTT AGCCGGCTGC TGCGAGCGTG CTTACGGGGA CCGCGGCCTG ACACCGTATC TTGCCTCTCC AACAGCCTTG GGGCGCGCGG CCCATGGAGT CGGGGCTGCT GCGGCCGGCG CCGGTGAGCG AGGTCATCGT CCTGCATTAC AACTACACCG GCAAGCTCCG CGGTGCGCGC TACCAGCCGG GTGCCGGCCT GCGCGCCGAC GCCGTGGTGT GCCTGGCGGT GTGCGCCTTC ATCGTGCTAG AGAATCTAGC CGTGTTGTTG GTGCTCGGAC GCCACCCGCG CTTCCACGCT CCCATGTTCC TGCTCCTGGG CAGCCTCACG TTGTCGGATC TGCTGGCAGG CGCCGCCTAC GCCGCCAACA TCCTACTGTC GGGGCCGCTC ACGCTGAAAC TGTCCCCCGC GCTCTGGTTC GCACGGGAGG GAGGCGTCTT CGTGGCACTC ACTGCGTCCG TGCTGAGCCT CCTGGCCATC GCGCTGGAGC GCAGCCTCAC CATGGCGCGC AGGGGGCCCG CGCCCGTCTC CAGTCGGGGG CGCACGCTGG CGATGGCAGC CGCGGCCTGG GGCGTGTCGC TGCTCCTCGG GCTCCTGCCA GCGCTGGGCT GGAATTGCCT GGGTCGCCTG GACGCTTGCT CCACTGTCTT GCCGCTCTAC GCCAAGGCCT ACGTGCTCTT CTGCGTGCTC GCCTTCGTGG GCATCCTGGC CGCTATCTGT GCACTCTACG CGCGCATCTA CTGCCAGGTA CGCGCCAACG CGCGGCGCCT GCCGGCACGG CCCGGGACTG CGGGGACCAC CTCGACCCGG GCGCGTCGCA AGCCGCGCTC GCTGGCCTTG CTGCGCACGC TCAGCGTGGT GCTCCTGGCC TTTGTGGCAT GTTGGGGCCC CCTCTTCCTG CTGCTGTTGC TCGACGTGGC GTGCCCGGCG CGCACCTGTC CTGTACTCCT GCAGGCCGAT CCCTTCCTGG GACTGGCCAT GGCCAACTCA CTTCTGAACC CCATCATCTA CACGCTCACC AACCGCGACC TGCGCCACGC GCTCCTGCGC CTGGTCTGCT GCGGACGCCA CTCCTGCGGC AGAGACCCGA GTGGCTCCCA GCAGTCGGCG AGCGCGGCTG AGGCTTCCGG GGGCCTGCGC CGCTGCCTGC CCCCGGGCCT TGATGGGAGC TTCAGCGGCT CGGAGCGCTC ATCGCCCCAG CGCGACGGGC TGGACACCAG CGGCTCCACA GGCAGCCCCG GTGCACCCAC AGCCGCCCGG ACTCTGGTAT CAGAACCGGC TGCAGACTGACACCCTCGGC CCACG	Sphingo-sine 1-phosphate receptor EDG-8 (SEQ ID NO:24)	P21453	Probable G protein-coupled receptor EDG-1	381	23	375	371	46

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
189901 LG895	AC013396	Genomic Clone	1218	GCCCTTACCC CCACAGCGCT GCAGCCCTGC AGCTGGCCCT CAGCCCTGGG AGGAGCCTTC CTTTTCCAGA GAGACCTCGC CCTGCACTTT CAGCTTCCCT ATGGCCTCCG CCTTCCTAGA GGCCTCCCGG TAGCGCCACT GCCTGGAGGG TTGGTAGGAG CTCTCGTCGC TCACTGGGCC CTGCCGGCCC CGCGTGAGGC CCAGCAAGGC CCGGCTCTGG TGGAGGAAGT TGGGGCTAGA GAAGCAGTAG AGCACGGGGT CCAGGACACT GTTGAGGTAG GTGAAGGCCA GGGAGCCATG GAAGAGCTGT GTGCAGAGGT CCAGGGATCG GCAGGCGGAC AGCCAGAAAG CCACCATGGA AGCCATGCCA AAGATGATGC TGGGCAAGAA GCAGATGGTG TAGACGGCCA CCACCATGGC CAGCACACGC ATGGCCCTCT GCGGGCCTGC CTGCCCGCCC AGACCACGGT TCCGGATGGT GAGCCCAATG CTCACAATAG CAAAGAGGAT GAGCGCCAGT GGCAGGAAGA ACTCCAGCAG GTACAGTGCC TGGTGCCAGC GGAGCGAGGC CGAGGGCTTC GTGCCCACCC TGTAGCTGAG GCAGGAGGGG CCGGAGAAGG TGCTCAGGAG CAGGTGCCCG TTGAGGAGCA GGATGCCCAC CCAGAGTCCC CCGGCCACCC GGGCAGCTGC CCCCACGGAA GCACGGCTCA GCACGTGGTG GGGCTGCACC ACCTTCAGGT AGCGGTTGAG TGCGATGGCT GTGAGGAAGA CAACGCTGGC CGTGCGGTTG GTGGACAGCA TGAAGAGGTT GACTTTGCAG GCAGCAGCCC CAAAGCGCCA GGTCTCATGG AGGAGGTAGT AGTCCACGCG GAGGGGCAGG TTGCTGATCA GGAGGAAGTC AGCGGCCACC AGGCTGACCA GGAACACCGT GTTGGAGGTC CAGGGCCGCG TGTGGATGCA GAAGATGAAG AGGGCCAAAC TGTTCCCCAC CAGGCCCAGG ACAAACTCCA GGGCCAGGAT TGGTGCCAGG AAGGCAGACA CCAGCGAGGA AGAGGTGGGG TGGCAGGGCC CTCCAGAGGA CCCCCCCACA GTGGTAAAGG CAGAGGGAGC AGAGGAGGGT GAGGGAGAGA AGGAGGGAGG GAGAACAGAG GAGGAGAGAG AGGGAGATGG AGAGCTCAGG TTATGAAGTT CCATGGGCTG CTTGGGCCAT GGGCCTGA	SEQ ID NO: 25	P49019	Probable G protein-coupled receptor HM74	387	18	301	280	43

LS Cluster ID: Current (Original) LG NO.	Acc. No.	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
190188 LG5982	AC018896	Genomic Clone	1005	ACAGGCCCCT TCAAGCCCTG TGAGTACCTC TTTGAAAGCT GGGGCATCCG CCTGGCCGTG TGGGCCATCG TGTTGCTCTC CGTGCTCTGC AATGGACTGG TGCTGCTGAC CGTGTTCGCT GGCGGGCCTG TCCCCCTGCC CCCGGTCAAG TTTGTGGTAG GTGCGATTGC AGGCGCCAAC ACCTTGACTG GCATTTCCTG TGGCCTTCTA GCCTCAGTCG ATGCCCTGAC CTTTGGTCAG TTCTCTGAGT ACGGAGCCCG CTGGGAGACG GGGCTAGGCT GCCGGGCCAC TGGCTTCCTG GCAGTACTTG GGTCGGAGGC ATCGGTGCTG CTGCTCACTC TGGCCGCAGT GCAGTGCAGC GTCTCCGTCT CCTGTGTCCG GGCCTATGGG AAGTCCCCCT CCCTGGGCAG CGTTCGAGCA GGGGTCCTAG GCTGCCTGGC ACTGGCAGGG CTGGCCGCCG CGCTGCCCCT GGCCTCAGTG GGAGAATACG GGGCCTCCCC ACTCTGCCTG CCCTACGCGC CACCTGAGGG TCAGCCAGCA GCCCTGGGCT TCACCGTGGC CCTGGTGATG ATGAACTCCT TCTGTTTCCT GGTCGTGGCC GGTGCCTACA TCAAACTGTA CTGTGACCTG CCGCGGGGCG ACTTTGAGGC CGTGTGGGAC TGCGCCATGG TGAGGCACGT GGCCTGGCTC ATCTTCGCAG ACGGGCTCCT CTACTGTCCC GTGGCCTTCC TCAGCTTTGC CTCCATGCTG GGCCTCTTCC CTGTCACGCC CGAGGCCGTC AAGTCTGTCC TGCTGGTGGT GCTGCCCCTG CCTGCCTGCC TCAACCCACT GCTGTACCTG CTCTTCAACC CCCACTTCCG GGATGACCTT CGGCGGCTTC GGCCCCGCGC AGGGGACTCA GGGCCCCTAG CCTATGCTGC GGCCGGGGAG CTGGAGAAGA GCTCCTGTGA TTCTACCCAG GCCCTGGTAG CCTTCTCTGA TGTGGATCTC ATTCTGGAAG CTTCT	SEQ ID NO:26	O75473	Orphan G protein-coupled receptor HG38	907	546	872	325	50
190408 LG5392	AC008969	Genomic clone	813	AACATTTGGG GCAAAGATGC CACCCAGCAG CCCTGCCATG GAAGCCAGGA TGGAAAAGAT CTCCACGGCC ACAGTGGACT TGCCCTGTGC GCTGTGGTAC AGGGGCAGGA AGGTTGTCCA GAAGCTGCAG AACAGCAGCA CGCTGAAGGT GAGGAACTTG GACTTGTTGA AGGCGTCTGG CAGACCCCTG GCCAGGAAGG CTACAAAGAA GGTGCCCCCA GCCAGGAGGC CCAGGTAGCC CAGCACACAG GAGAAAGCGA CAGCAGAGCC CTCTCAGCAC TGGATGACAA TGTGGCTGGG CTCTGAGGCC ATGTCCCCAT CTGGGAATGG TGGGGAAGTG CCCAGCCAGA TGCCACAGAG AACAACCTGC ACCAAGGAAG CAGCAAGGAC CACCGAGGTG GAAGCGCCAG GTCCCAGGCA CACCCAGACC CTGTCACCTG GTGACCCTGA AGGCCAGGAC GGAAGAGACA GCCATGGTGA ACACAACAGC AACTGTGGTC TGGTAGAGGC GGCAGGTGGC AGCTGTGGGA CAACCAAGGC AAGGCAAGGG ACAGAGGGCA CAGAGGGTCA GGGAGGTGAG CGGCGTGCAG CTGAGAGCTC TGTTGTTGGC CCCGACCACA GGTGTGTCTC GGTGCTTCAG AACAGTCTGA GGACCAGCAC TGCCAGGCTG GCCAGCATGA GCGCCACCAA GGCGAGCATG AGTCCCAGGG GGTCGTCAAA GGCCAGGAAG GTCTCTGTCC TGGGCAGGCA GCCATCTCTG GTGCGGCTTG AGTACTGCTC CTCTGGGCAC AGAAGACATC TCTTCGTACC TGT		P41594	Metabotropic glutamate receptor 5 precursor	1212	744	822	81	35

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
190411 LG5580	AC011457	Genomic clone	769	AATAGTGTTC CTTTCTAGAT GAAGGAGGAT CACATAGCAC TTGGGAGCAA AGATGCAGCC AAGTAACCCA GTGCTGGAGC CCACGATGGA GAAGATCTCA CGGCCACTCT GGCCTTGCCC TGGGTGCTTT AGTAACTCGG GAGGAAGGCC ACCCAGACAC TGCAGGACAC CAGCATGCTG AAGGTCAGGA ACTTGGCTTC GTTGAAAGCA TCAGGCAGGT TCCTTGCCAG AAAGGCTACA GCAAGGGACC CTAAAACCAA GAAGCCCAAG TAGCCCAAGA CAGAGTAGAA GGCAGTGACG GAGCCCTCAT TACACTGGAT AATGATGTAG CCAGGCATGA ACTGAGGGTC CTTGTTTACG AAGGGAGGCT CTGTCCCCAG CCAGATTCCA CAGAGGGTCA CTTGGGTAAA GGAGCAAAGG AAGACAACAG AATTTGTTTT TTGGGGACCC ATCCACTTTC GGATCCTGCT TTCTGGCCTT ATAGCCTGGA AGGCCCTGCA GAAAACACAG CCACTGTGAA CACAGCTGCA AAGGTGGTCT GTGAGAGGAG GCAGGTGGCA GGGCTGGGGC GGCCGATGAA GAGCGAGGAA GAGAGAAAGC AGAGTGTGAG GGAGACAAGG AGGGTGTAGC TCAGAGTCTG GTTGCTGGCC TTGACTATGG GAGTGTTTTG GTGCCACACA AACTCCACAA GGATCAGAAC AGAGAGGACA GAGTAGAACA AGGCTGTGCA TGCCAGAGTC AGGCCCAAAG TTTCATCATA GGCCAGGAA	Phero-mone receptor SEQ ID NO:27	P41180	Extracellular calcium-sensing receptor precursor	1078	706	850	145	42
190412 LG5459	AC010136	Genomic clone	652	CAGAAATCCT CAGGTCCCAC AGAAATGAAC ACGTTTTCTA AAATAAAGTC AAGCCAAGCT GTCCTACCCC AAAGAAAATC CTAGCAAGCA AAGGTGGCTT CCTTCCTGAG GCCCCAGCCA GGTGTGTCCA ACCGTAGGAG CCACAGCTCA GAGATCAGAG TGACTTAACA GTTAGAGGGC ACTTGATGAG TAAGGTGAAA TAGGGAAACC AAGTCAGACG ACACCTCCCT TCTGAGTCCC AACCATGTCT ACATCTGGAG AAGAACAGTT AAGTCAAGGG ATCACAGACT TGTGATTAGA GACTGCCAGG GTCCATATGA CCAAGCGGGG GTCCCAGGTG TGAAGCTGGG GTTGAGGATC CATTATCTGA ATTTTCCACT CTATGGATGA TCACTTTTAT TCTTTTCCTT TTCTTGAATT TATTTCCATT TGTTTATTCC TAAATTCCCT GGTAGATCAC CTGTGAAAGC TTGCAACTGT CTGATAAGAA TAAAGGGGGA AGGATTTGAC TTTACAGCAG AGACTTCAGA AGGAGTCCTC TCTAGGAGCA AATTGGGGGC AATCCAGTGG GAAGGAGGTG GAAGACTGCA CTTGAGCTGC GTTTGGACAA CAGGCACACA ATCTTTACTT ACTTTTCAGG CTGCTTTGAG GT		L19592	Interleukin 8 receptor alpha (IL8RA)	9269 bp	3917	3266	654	89

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
190414 LG5853	AC016468	Genomic clone	1575	CTTTTGTCAC CAGCATAGGC ACTGAGTGCG GTCTGTGCAC CCCTTTGCCA CCCACCGGTG CCGGCACTGA GCCTGCAACC TGTCTCACGC CCTCTGGCTG TTGCCATGAC GTCCACCTGC ACCAACAGCA CGCGCGAGAG TAACAGCAGC CACACGTGCA TGCCCCTCTC CAAAATGCCC ATCAGCCTGG CCCACGGCAT CATCCGCTCA ACCGTGCTGG TTATCTTCCT CGCCGCCTCT TTCGTCGGCA ACATAGTGCT GGCGCTAGTG TTGCAGCGCA AGCCGCAGCT GCTGCAGGTG ACCAACCGTT TTATCTTTAA CCTCCTCGTC ACCGACCTGC TGCAGATTTC GCTCGTGGCC CCCTGGGTGG TGGCCACCTC TGTGCCTCTC TTCTGGCCCC TCAACAGCCA CTTCTGCACG GCCCTGGTTA GCCTCACCCA CCTGTTCGCC TTCGCCAGCG TCAACACCAT TGTCGTGGTG TCAGTGGATC GCTACTTGTC CATCATCCAC CCTCTCTCCT ACCCGTCCAA GATGACCCAG CGCCGCGGTT ACCTGCTCCT CTATGGCACC TGGATTGTGG CCATCCTGCA GAGCACTCCT CCACTCTACG GCTGGGGCCA GGCTGCCTTT GATGAGCGCA ATGCTCTCTG CTCCATGATC TGGGGGGCCA GCCCCAGCTA CACTATTCTC AGCGTGGTGT CCTTCATCGT CATTCCACTG ATTGTCATGA TTGCCTGCTA CTCCGTGGTG TTCTGTGCAG CCCGGAGGCA GCATGCTCTG CTGTACAATG TCAAGAGACA CAGCTTGGAA GTGCGAGTCA AGGACTGTGT GGAGAATGAG GATGAAGAGG GAGCAGAGAA GAAGGAGGAG TTCCAGGATG AGAGTGAGTT TCGCCGCCAG CATGAAGGTG AGGTCAAGGC CAAGGAGGGC AGAATGGAAG CCAAGGACGG CAGCCTGAAG GCCAAGGAAG GAAGCACGGG GACCAGTGAG AGTAGTGTAG AGGCCAGGGG CAGCGAGGAG GTCAGAGAGA GCAGCACGGT GGCCAGCGAC GGCAGCATGG AGGGTAAGGA AGGCAGCACC AAAGTTGAGG AGAACAGCAT GAAGGCAGAC AAGGGTCGCA CAGAGGTCAA CCAGTGCAGC ATTGACTTGG GTGAAGATGA CATGGAGTTT GGTGAAGACG ACATCAATTT CAGTGAGGAT GACGTCGAGG CAGTGAACAT CCCGGAGAGC CTCCCACCCA GTCGTCGTAA CAGCAACAGC AACCCTCCTC TGCCCAGGTG CTACCAGTGC AAAGCTGCTA AAGTGATCTT CATCATCATT TTCTCCTATG TGCTATCCCT GGGGCCCTAC TGCTTTTTAG CAGTCCTGGC CGTGTGGGTG GATGTCGAAA CCCAGGTACC CCAGTGGGTG ATCACCATAA TCATCTGGCT TTTCTTCCTG CAGTGCTGCA TCCACCCCTA TGTCTATGGC TACATGCACA AGACCATTAA GAAGGAAATC CAGGACATGC TGAAGAAGTT CTTCTGCAAG GAAAAGCCCC CGAAAGAAGA TAGCCACCCA GACCTGCCCG GAACA	SEQ ID NO:28	P08912	Muscarinic acetylcholine receptor M5	532	2	518	470	23

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len	From	To	Aligned	Percent
189886 (190416) LG6804	AL136961	Dbest	1001	TGGGGTAATG CCACAGTTCA TGTTCATGAG GGCCACGGTG GCCTGAAGGG ACAGTAAGAA AGCCCTCCGC TCGGCACAGG ATGGCAGGTG GAGCATCCCT CTCGCCATGA ACTGCTTGAT GTTGAGGTGG TAGGGGCTGA AGCAGACCAC CACGGCCACC AGCATCAGCA GCGTAAGCAG GCAGCCTCGC CAGTGGCGTC CTTTCCTGCT GGTCACTGGG TCCTCCCGGG CTGTCCCGCC GGTGGCGTCC TTTCCCGCTG GTCACTGGGT CCTCTCGGGC TGTCCTGCCG GTGGTGTCCT TTCCCGCTGG TCACTGGGTC CTCCTGGGCT GCCCCGCCGG TGGCGTCCTT TCCTGTTGGT CACTGGGTCC TCCTGGGCTG CCCCGCCGGT GGTGTCCTTT CCCGCTGGTC ACTGGGTTCT CCCAGCTGTG CTGCACAGCT TCCAGGTGAT CTTCATATAG CAGGACAGGA TGATCCCCAC TGGCCCACAG AAGCCAATGG CAAAGGCCAC CAGGACCATG AGGGGCAGCC CGAGGACTGA CTCCATGCTG CTGTACTCCA TGCAGGCCAG CTTGCCCACC AGCGGCTTGG TCATGGGCAT CAAGAGCAAG GGCATCGTCT GCAGCAGCAC CAAGGTCCAG ATGGCCACGC AGACCAGCCT GGCGCGGCCA GCCGTGCGGA GGCGCGGGCC CCAGTGGGCA CAGACCACAG CTGGGTAATG GTCCACGCTC ACACAGGCCA TGAGGTAGAC CCCCCGTAGG TGTCGGTGTA GAGCACAAAC GCCGTCAGCC TGCAGAGCCC CTTGCCGAAA GGCCAGCTGG AGCCCAGCAC ATAACACACC ACCCTTCCCG GTAAGGCCAC GGTGAACAGC AGGTCAGACA CAGCCAGGTG CACCAGGTAG ATGCCTGTGC AGTTGATCTT CCTGCTCTTT TGACAGGTAA GGCAAAGGGC AAGGATGTTT CCCAGGGCAC TGAAGACCAG GAGGGCTGTG TAGAACAGAG A	SEQ ID NO:20	P32249	EBV-induced G protein-coupled receptor 2	361	36	118	83	44

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
189889 (190417) LG5881	AC016856	Genomic clone	947	TGCTTGACCA CGGAAAAATT TCCAAAAGCT TCCCCACTCT GTTGTCTGAC CCAGTTTCTC TTTCTCCCTT CAGATAGTGT TGCATGTCAG AAGTGCTCTG ACAACCAGTG GCCCAATGTG CAGAAGGGCG AGTGCATCCC CAAAACCCTT GACTTCTTGT TCTATCACAA GCCCCTTGAC ACAGCGTTGG CTGTCTGCAC AGCCCTGCTC TTTCTCCTTG CCCTGGCCAT CTTAGGCATC TTCGTCTGAC ACCACCACAC TCCCATCATC CGAGCCAACA ACTGCCAGCT CAGCTATCTC CTGCTGTCCT CCTTGGCCCT CAGCTTCCTC TGCCCCTTCA TGTTCATTGG CCACCCAGAC CCCATCACTT GTGCTGTGCA CCAGGCAGAT TTTGGGGTCA CCTTCATGGT CTGCACATCC ACTGTGCTGG CCAAGACCAT CGTGGTGGTG GCAGCCTTCC ATGCCACCCA GGCAGACACC CAGCTTAGGG GGTGGGCGGG GACAGTCCTC CTCAGCACCA TCCTCACTGT TCCCTGACCC AGGCAGCCTT GTGTGCACTC TGGGTGACCA GATGGCCCCC TCAGCCTGTG AAACTCTACA GAACCCTGGC CCACAGTGAC TGTAAAGTGT GATAAAGGCT CCTTGGAACT TCTCTTGGAA CTGGGCTACT TGAGTTTGCT AGATCTGGTC AGCTTGCTGG TGACCTTCCC CACCTGCCGG CTGCCTGACA CCTTCAATGA AGCAAAGCAT ATCACTCTCA GCATGTTGTC TGCTCCTGTG TCTGGGTGTC CTTCATACCT GCCCACATGC ATGCCCACAG CAAAGACACC ATGGCCATGG AGGTCTTTGT CATCTTGGCA TCAGCAGGAG GCCTCATGTC CTCCCTCTTC TTTTCCAAAT GCTACATCAT CCTTCTCCAT CCTGAAAAGA ACACAAAAGA CCAAATG	SEQ ID NO:21	P41180	Extracellular calcium-sensing receptor precursor	1078	577	828	246	36
190418 LG6080	AC020641	Genomic clone	840	TTTTAAAATG GAGCCATATG CTTGGCGGAA TTGGCGGTTC ATGGCTGCAT AGAGCACAGG GTTGATGCAA CCATTGAGCC AGGTGAGGTT GGCAGCAAGC ATGTGGACCA CCCGGGGAGC CTGGACTCTG GCATCCAGAA TGTTGAGCAG CAAGAAGGGG ATGTAGCTCA GGGCAAAGCA GAGGAACACA GCAAAACACA TTCGAGTCAC CTTCCCAAAT TCCGATGAAG AATCCGGAGC TCTTCTGGCT CCTTTAATTG GCTGGGCTTT GGCAGATGCT TCTGGAGGGC TTTTCTCTGC CATCTGCTTA GCTCTCTTGC TGTTGATCTG GTCTCCCACT TCTGATGAGT CCCCTTCCAG GGTCTGGGTG GTGGCAGCAC TGACTGGCTC AGATGAAATC CCCTCACTGG GTCCTCCTGA TGCTAACCTG CTGTCCAGCT CCTGGAAACG ACCAGGCATG GCCTCATCAG TCCTGGCCAC ATGGTTGGAG TGGATGCTTG CCTGTCGCAA CTTGTATTGG TCCAGTGCCT GTGCTGCTCG TTTGACCTGG CGGTGGATGA GGCAATAGAA GATGCCAACA CTGCTGAGCC CAAGCACAAA GTAGATGCCC ATGAGGATGG TGGTGTAAGG CCGGCCTCGG ATGCGGTCAA AGCTGCAGGT GCAGACTACA GGTACCAGGA TATAAATAGG CCAGAGGGGA GCAAAGCTGG CCACGCCCAC AACCCAGGTG CTCACCAGTG CCAGCACTAT CCCCTTGGCA CTGAAAACTT GGGGAAAAAG CTTAGGGTGG GCAATGAGGA GGTAGCGTCC CAGTGCGATG AGGCAGAGGG TCAGGATGGA	GPR 84 EX33 SEQ ID NO:29	P28566	5-hydroxy-tryptamine 1E receptor	365	109	360	280	25

LS Cluster ID: Current (Original) LG NO.	Acc. No.	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
190419 LG6171	AC021089	Genomic clone	867	CTTTGCTTCA GAGCTAAACC AGTTTTTCTT CTCTCCACAG CAAATATCTT GACAGTGATC ATCCTCTCCC AGCTGGTGGC AAGAAGACAG AAGTCCTCCT ACAACTATCT CTTGGCACTC GCTGCTGCCG ACATCTTGGT CCTCTTTTTC ATAGTGTTTG TGGACTTCCT GTTGGAAGAT TTCATCTTGA ACATGCAGAT GCCTCAGGTC CCCGACAAGA TCATAGAAGT GCTGGAATTC TCATCCATCC ACACCTCCAT ATGGATTACT GTACCGTTAA CCATTGACAG GTATATCGCT GTCTGCCACC CGCTCAAGTA CCACACGGTC TCATACCCAG CCCGCACCCG GAAAGTCATT GTAAGTGTTT ACATCACCTG CTTCCTGACC AGCATCCCCT ATTACTGGTG GCCCAACATC TGGACTGAAG ACTACATCAG CACCTCTGTG CATCACGTCC TCATCTGGAT CCACTGCTTC ACCGTCTACC TGGTGCCCTG CTCCATCTTC TTCATCTTGA ACTCAATCAT TGTGTACAAG CTCAGGAGGA AGAGCAATTT TCGTCTCCGT GGCTACTCCA CGGGGAAGAC CACCGCCATC TTGTTCACCA TTACCTCCAT CTTTGCCACA CTTTGGGCCC CCCGCATCAT CATGATTCTT TACCACCTCT ATGGGGCGCC CATCCAGAAC CGCTGGCTGG TACACATCAT GTCCGACATT GCCAACATGC TAGCCCTTCT GAACACAGCC ATCAACTTCT TCCTCTACTG CTTCATCAGC AAGCGGTTCC GCACCATGGC AGCCGCCACG CTCAAGGCTT TCTTCAAGTG CCAGAAGCAA CCTGTACAGT TCTACACCAA TCATAAC	SEQ ID NO:30	O14694	CCR5 receptor (fragment)	333	22	304	270	27

ES Cluster ID: Current (Original) LG NO.	Acc. No.	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
190420 LG6269	AC021773	Genomic clone	1063	CCTTCCAGGG GAAGCAGAAG CGGGAGCCGT CGTGGAGCTC TGCTCCTGGA GGGAGCCTCC CGGGACATGG AGAAGGTGGA CATGAATACA TCACAGGAAC AAGGTCTCTG CCAGTTCTCA GAGAAGTACA AGCAAGTCTA CCTCTCCCTG GCCTACAGTA TCATCTTTAT CCTAGGGCTG CCACTAAATG GCACTGTCTT GTGGCACTCC TGGGGCCAAA CCAAGCGCTG GAGCTGTGCC ACCACCTATC TGGTGAACCT GATGGTGGCC GACCTGCTTT ATGTGCTATT GCCCTTCCTC ATCATCACCT ACTCACTAGA TGACAGGTGG CCCTTCGGGG AGCTGCTCTG CAAGCTGGTG CACTTCCTGT TCTATATCAA CCTTTACGGC AGCATCCTGC TGCTGACCTG CATCTCTGTG CACCAGTTCC TAGGTGTGTG CCACCCACTG TGTTCGCTGC CCTACCGGAC CCGCAGGCAT GCCTGGCTGG GCACCAGCAC CACCTGGGCC CTGGTGGTCC TCCAGCTGCT GCCCACACTG GCCTTCTCCC ACACGGACTA CATCAATGGC CAGATGATCT GGTATGACAT GACCAGCCAA GAGAATTTTG ATCGGCTTTT TGCCTACGGC ATAGTTCTGA CATTGTCTGG CTTTCTTTCC CTCCTTGGTC ATTTTGGTGT GCTATTCACT GATGGTCAGG AGCCTGATCA AGCCAGAGGA GAACCTCATG AGGACAGGCA ACACAGCCCG AGCCAGGTCC ATCCGGACCA TCCTACTGGT GTGTGGCCTC TTCACCCTCT GTTTTGTGCC CTTCCATATC ACTCGCTCCT TCTACCTCAC CATCTGCTTT CTGCTTTCTC AGGACTGCCA GCTCTTGATG GCAGCCCAGT GTGGCCTACA AGATATGGAG GCCTCTGGTG AGTGTGAGCA GCTGCCTCAA CCCAGTCCTG TACTTTCTTT CAAGGGGGGC AAAAATAGAG TCAGGCTCCT CCAGAAACTG AGGCAGAACA AGTTGGGTGA GCATCCAGCT GGGAGGAAGA GATGCCCAGG GTTGAACAGA TCT		Q15077	P2Y purinoceptor 6	328	4	202	193	43
190421 LG6465	AC023078	Genomic clone	729	AAAGNACCCC CACGTGGNAT CCATAAAAAT AGGGAAAAAC TGAATGCCAA AGGGCAGGCC ACAGAGGAGG AAGACCAGTA CTGTGAGCAG GATGGTCACG TACAGCCTGG TCAGCGGTAT CTTCCGGGAT CCACAGAGAA TCCTGATCAG CAGGACCAGG CTGGACCCAC AGAGAACCAC ACATAAAAAA ATCAGCCACG CGACTGTGAT GAAATCTGAT GTTTGACACC AAGCAGAATC AGCACCACTG AACAGGAAGC CACATAACAT CCACTCCAGG ATGCTCCGCA GCAGGGACAG GGCCCAGAGC AGGACACACA CCACCGCTGA CAGGTGTGTG GGGCGGTGGC AGCGGTACCA GATGGGCCAC AGGACGGACA GGCAGCGCTC GGTGCTCACG GCACTCAGAA AGCTCAGGCC TGCAAAGTAG GAAAACATCA TCACAGGATA GAGGATTTTA GAGATGGTAT GGGGGATACT GATGAAGCTT AACAGGGAAT ATATAAGGCG GCCGCTGAGG AAGAGGAAGT CTGCTGCGGC CAAGTTGAGG ATGTAGATGG AGAAGGCGTT CCTGCGCATG CGGCAGCCCA GGAGCCAGAG CACAACTGCG TTTCCTGTCA GCCCGACAAG GGAAACGATG CACGTCAGCA CCGTGAGGCT CAAGGTCTGC TTGTAGCAAA GAGTCTCCTC AGTTCCGTTG ATTGGTGTCA GTTCTGTGTC CAAGGTTGA		P35410	MAS-related G protein-coupled receptor MRG	378	70	282	210	38

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
190423 LG6564	AC023497	Genomic clone	461	CCGGCGCCAC GTGCCTGCTG CTGCGCGCCT ACGTGACGCG GCATTGTCAT GCACTGGCTG ACCTATCATG AGACCCTGCT GCTGCTCACA CTGTATGGAA CCCACATCTG CCTACACTGC CACCTGGTAC CAACTGCTCT ACTTCTTCTA TGATGTCATC TGACTGCTGC TACATGCTAG ACTGCGCTAT TCACCGGATC CTTGACAACT TTATCAGCCA GACTGCCGGG GGCGGCTGCG ATGCTGTGGT CCATTACTTG CTAAGGACCA GACCGCGGGG GCACATGCGC CTCCTCTTCC TTCTGTGACA CCCAGCGTTA CATAATCATT ACCACGGGTG ATAGCCAGAC TGCTGCGAGC AACCCGCCAC CCTGCAGCCA AGCCTGAGCT TTCAGGCACA CCATTCGCTC GCAAAGACTT GCGCCATGTG TCCCACTCAG TGTCTTACAC CCAGCTGAGG T		O15218	G protein-coupled receptor	404	268	377	109	45
190424 LG6770	AL133460	Genomic clone	385	TGCCAAATAT GCTGTTGCCA ACACTTAGAA CACAATGACT GGAGACACAG TTGTGCGTGC CTGGCACAAC CTCCAGCCTG TGTCTATGTT CAGTGATGAT GATGAGCAAG GTGGTGACTT TGAAGGATTT TGTATATCAA GTGAAAAGAA ATGATATCTG ACCTCCTTAC ATATCTAAAA CATATACCTT CAAAATCCAT CAATAAGCTG AAAGAAATAG ATATCAAAGA ATATTTTAAC ATCATTAATG AGGCTCCAGT TATTCATTCA TTGACCAATG GTAATATAGC TGAAATGATT CTGAATCAAG CTGATTATGA TAATAGTGAT GATGAAGATG ATGTTAATAC TGCAGAAAAA GTGCCTATAA ATGACACAGT GAAAA		U62556	Chemokine receptor-like protein (TER1)	2608 bp	1536	1915	388	81
190425 LG6786	AL136106	Genomic clone	429	GCTGATTGCC TGATGGGTGT TTACTTGTTC TTTGTTGGCA TTTTCGATAT AAAATACCGA GGGCAGTATC AGAAGTATGC CTTGCTGTGG ATGGAGAGCG TGCAGTGCCG CCTCATGGGG TTCCTGGCCA TGCTGTCCAC CGAAGTCTCT GTTCTGCTAC TGACCTACTT GACTTTGGAG AAGTTCCTGG TCATTGTCTT CCCCTTCAGT AACATTCGAC CTGGAAAACG GCAGACCTCA GTCATCCTCA TTTGCATCTG GATGGCGGGA TTTTTAATAG CTGTAATTCC ATTTTGGAAT AAGGATTATT TTGGAAACTT TTATGGGAAA AATGGAGTAT GTTTCCCACT TTATTATGAC CAAACAGAAG ATATTGGAAG CAAAGGGTAT TCTCTTGGAA TTTTCCTAGG TAAATTATAT TTTTTCATT		Q15996	Luteinizing hormone receptor	699	404	541	143	31

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
190427 LG6807	AL137118	Genomic clone	1026	TTCCTTTCTC AACCACACAC TAACAGGGAA AACACACTTT GTCTTTGCCT TCTGTGGATG GCCTTTTCTG AGTGCAGACT TTAGTCTGTC CTTAAAATTC TCCCCAGCAA AGTAATAGAG CAGAGGATTG AAGCAGGCAT TGGCTGCTGC CAAGGCCAGT GTGATAACCA AAGCTTTATG CAGTCTGTCT TTGCATAAAC CCACTTTCCA TGTCGTCAAG TGGACGGTCC TCAGTGTGTG ATAGGGCAGG AAACACAAGA AGAAGATGAT CAAGGTGATG ATGATGGTGG TCAGTGCCTT CCTGTGAGAA ACCCGCAGCC CCGATTCTGG GACCTCCACT TTTAACAGAA CCCGAATGAT CAGCAGATAA CAGATGCTGA GTGTGAAAAA TGGCAGCAGG CAGCCCACCA CCAAGGCAAT ATAGTTCATG GTCTGCAGCT TAGCAATTTT ATAGAGATTC AGCTCTAAGC ATGATGTGAC ACTGCCGTTC TGCTCAGAGC CACTGTCCAG GAGCATTATT GAGGAAGCCA TGATAAGGAT CCATATGATC CCACAGAGGA TCCAGGCACT CCTGATGCTG GTGACATGCA GAAGCCGAAA GGGGTGAACC ATTGCCAGGA AACGCACAAC ACTCAGCACG GTCAGGAAAT AAATACTGCT GTACATGTTG ACATACAAGG AATAAGACAT AATCCTGCAG GCCAGGTCTC CAAATATCCA ATTGGAGCCT CTAAGATAAT AGTCAGCCCT GAAGGGAAGC GTGCTTATGA ACAGGAGATC TGAAATGGCC AGATTTAGCA TGAAAACGTT CACAGATGTG GACTTCTTAT AAGGCTGCAG GAAAACATAT ATGGACAACC CATTTCCCAA GACTCCCCAG AAAAATATTA TCAGATATAC AATTGGGAAA AATTCTCTCT TGAAGTTTTC AATTGTGCAG TTCCTGCTGT TGTTATTGCT GAAGGTGCCA TTTGGTTCCA TTTCTGATAC GGAGATGGAT GGTTGCAAGG ACATAAATTT TCTCTC	Cystein-yl leuko-triene CysLT2 receptor SEQ ID NO:31	Q9Y271	Cysteinyl leukotriene receptor 1	337	17	311	291	39
190428 LG6894	AP000440	Genomic clone	426	AGAGTCATCT GCCTCATTAA TGATGCTTTT TGTCTTAGAA GTCTCTCTTT GATTTTATAA ACTGACATGA TTTCTTGCTC TGTTATGAAT GCATGCTGCT CTAGTCCTTT AATAAGCCCA TCACACATTT TTACCATGTC GTCTATAGGC ACTTTTTCTG CAGTGTTAAC ATCATCTTCA TTATCATTAT CATCATGATC ATCTTGATTC AGAACCATTT TTGTTACTTC ACCATTGGAC AATGAATGAA CAACTGGAGC CTCATTATTA CTGTTAAAAA CATCTTTGAT ATCCTCTTCT TCCAGCTTAC TGACAGTCTC TGGATGTATA TTTTTTGTAT ATGTAAGGAG GTCAGACATT ATTTTTTTCT CACTTGACAT ACTGAATCCT TCAAAGTTAC CATCTTGTTC ATCATCATTA CTGAAC	SEQ ID NO:32	U45983	CCR8 chemokine receptor (CMKBR8)	1944 bp	1941	1586	362	83

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
190430 LG5259	AC005883	Genomic clone	549	ATCAATGAAG AGTAAATATT AAAACACTTT CTTGAACAGC ACGCTAACAG CAGTGAAGAG AGGCATGGGT GTGACAGGTT TTCAAAAGGA GTTTAGCATG AAGGATGCCA TATATGCTGT TGCCAATACT TGGAATACTT GGAACACAGT GACTAAAGAC ACAGTTGAGC ACGCCTGGCA CAAGCTCTGG CCTGCAACTA CATTCAGTGA TGATGATGAA CAAGGTGGTG ACTTTGAAGG ATTCTGTATG TCAAGTGAGG AAAACATGAT GTCTAACCTC CTTACATATG CAAAAAACAT ACCTTCGGAG TCCATCAGTA AGCCGGAAGA AATGGATATC AAAGAAGTTT TTAACATTAA TAATAAGGTT ACAGTTGTTC ATTCATTGAC CGATGGTGAA ATAGCAAAAG TGATTGTAAA TCAAGGTGAT TGAGATAACA GTGACAAAGA TGACGTTGTT AACACTGCAG AAAAAGTGCC CATAGACAAC ATGGAGAAAA TATGTACTGG GCTTATTGAA GGACTAGAGC AGCATGTGTT CATAACAGA		U45983	CCR8 chemokine receptor (CMKBR8)	1944 bp	1328	1864	553	79
190419 (190431) LG5386	AC008785	Genomic clone	894	GTTATGATTG GTGTAGAACT GTACAGGTTG CTTCTGGCAC TTGAAGAAAG CCTTGAGCGT GGCGGCTGCC ATGGTGCGGA ACCGCTTGCT GATGAAGCAG TAGAGGAAGA AGTTGATGGC TGTGTTCAGA AGGGCTAGCA TGTTGGCAAT GTCGGACATG ATGTGTACCA GCCAGCGGTT CTGGATGGGC GCCCCATAGA GGTGGTAAAG AATCATGATG ATGCGGGGGG CCCAAAGTGT GGCAAAGATG GAGGTAATGG TGAACAAGAT GGCGGTGGTC TTCCCCGTGG AGTAGCCACG GAGACGAAAA TTGCTCTTCC TCCTGAGCTT GTACACAATG ATTGAGTTCA AGATGAAGAA GATGGAGCAG GGCACCAGGT AGACGGTGAA GCAGTGGATC CAGATGAGGA CGTGATGCAC AGAGGTGCTG ATGTAGTCTT CAGTCCAGAT GTTGGGCCAC CAGTAATAGG GGATGCTGGT CAGGAAGCAG GTGATGTAAA CACTTACAAT GACTTTCCGG GTGCGGGCTG GGTATGAGAC CGTGTGGTAC TTGAGCGGGT GGCAGACAGC GATATACCTG TCAATGGTTA ACGGTACAGT AATCCATATG GAGGTGTGGA TGGATGAGAA TTCCAGCACT TCTATGATCT TGTCGGGGAC CTGAGGCATC TGCATGTTCA AGATGAAATC TTCCAACAGG AAGTCCACAA ACACTATGAA AAAGAGGACC AAGATGTCGG CAGCAGCGAG TGCCAAGAGA TAGTTGTAGG AGGACTTCTG TCTTCTTGCC ACCAGCTGGG AGAGGATGAT CACTGTCAAG ATATTTGCTG TGGAGAGAAG AAAAACTGTT TAGCTCTGAA GCNAGATACT TCGTGCTCCT ATAGGCCGTA GGTC		O14708	CCR5 receptor (fragment)	352	24	323	286	26

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
190705 (190432) LG5394	AC008971	Genomic clone	816	GAATTGTTTT GGAATAAAAA TATATGCTAA GTAAGCCGGT ACCCCAGCAG CTTCTTCTGC GAGTGGTAGA GGCCCAGCCA GAACTGCCTG TCGCGGCCCA GCAACTTGTC CGGGAAACGC ACCAGTGCAC AGCTCCTCGC CCATCACCTT GACCGTGGTG GAGAAAATGG CACTGGGCAG CGAGGCCAGC GCGGCCAAAG CCCAGATCCA CACACACAGC GCCTTGGCCG AGAAGCAGCA GCTGTCCCCC AGGCTCCGGC CGCAGCAGTC GCCCCGGCCG TGTCCTCGGG TCCGGTGGCT CTTCAGAGCC GAGGCCACCG AATGGTAGCG CGTCACACTC ATGGCAGTGA GGAAGAACAC GCTGGCGTAC ATGTTCATGG ACGTCACCAT GGACACGATC TTACACATGG CCTTGCCGAA GGGCCATTTG AAGTCAAGAG CGTTCTCCAC CGCCCAGAAG GGCAGGGTGA GCACAAACTG AAAGTCCGTC AGCGCCAGGT TGGTGACGAA GAGGTTGATA GAGGACTTGC GCCAGCCCTG CATGCTCTTC ATCAGGTAGA GAACCAGCAG GTTGCCCGCC AACCCCAGGG CGCACACCAC CCAGTACACC ACGCTGATGA GAATCCGCAC CCGGGCCTCT GTGTCCGCGC TCTCTGCCCC GCCGCTGCCC GGGGGATGTC CTGGCGGCGC GCCGTCCGGC AACTCCAGCC CCAGCTCCCA CCACAAGTCC GGAAGCTGCA GCGACGCGTT ACCACTCGTG TTGGCCGCCT CCAGAAGGTC CGGGACCAGA CTGAAGAGTT CTGCTAGCTT GTCCCC	GPCR SALPR SEQ ID NO:35	P41143	Delta-type opioid receptor	372	6	203	215	33

LS Cluster ID: Current (Original) LG NO.	Acc. No.	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
160833 (190435) LG5416	AC009404	Genomic clone	1090	CCGGCGTTGC CCTTGCCCTG GGACCCGAGA CCAGCAGCAG GACCCGGGAC CCCAAGCCCG AGAGGGATAC TCGGTTCGAC CCCGAGCGGC GCCGTCCTGC CGGGCCGAGG GCCGCCCTTC TCTGTCTTCA CGGTCCTGGT GGTGACGCTG CTAGTGCTGC TGATCGCTGC CACTTTCCTG TGGAACCTGC TGGTTCCGGT CACCATCCCG CGGGTCCGTG CCTTCCACCG CGTGCCGCAT AACTTGGTGG CCTCGACGGC CGTCTCGGAC GAACTAGTGG CAGCGCTGGC GATGCCACCG AGCCTGGCGA GTGAGCTGTC GACCGGGCGA CGTCGGCTGC TGGGCCGGAG CCTGTGCCAC GTGTGGATCT CCTTCGACGC CGGAGCCTGT GCCACGTGTG GATCTCCTTC CACGGCTGTG CTGCCCCGCC GGCCTCGGGA ACGTGGCGGC CATCGCCCTG GGCCGCGACG GGGCCATCAC ACGGCACCTG CAGCACACGC TGCGCACCCG CAGCCGCGCC TCGTTGCTCA TGATCGCGCT CGCCCGGGTG CCGTCGGCGC TCATCGCCCT CGCGCCGCTG CTCTTTGGCC GGGGCGAGGT GTGCGACGCT CGGCTCCAGC GCTGCCAGGT GAGCCGGGAA CCCTCCTATG CCGCCTTCTC CACCCGCGGC GCCTTCCACC TGCCGCTTGG CGTGGTGCCG TTTGTCTACC GGAAGATCTA CGAGGCGGCC AAGTTTCGTT TCGGCCGCCG CCGGAGAGCT GTGCTGCCGT TGCCGGCCAC CATGCAGGTG AGGGGTGGGC TGAGGAACGT TGCTATGGGG AAGCGGTTGC TAGAGAAGGA GGCAGCTTCG CGAATGGGAG AGTGGGCGGA AGCTTGTACT AATGGAGCGC GCGCCCAGAG ATCACCTGGG GCGCACGAGG ACAAGTTTGC CATCAGTTCT TCTGAAGCGG GCACCGAGGG GCTTGTCACC GGCTCACCTG GCACGCAAGT GAGGGGTTCG CCAGCTGCGT ACTTGGTGCG CGCAGAGGAA AGGGTTTCAC AATCTGCACG TTAGGGAGGC AGATCTGACA CCGACCCACG ACGCACGATA	5-HT5B receptor (SEQ ID NO:11)	P47898	5-Hydroxy-tryptamine 5A receptor	357	123	257	134	55
189883 (190436) LG5393	AC008969	Genomic clone	792	CACCAGCCAG GGCTGGAAGT CCCAGGCATC GGCACACCCC TGGTCCCCAT GCTTGCAGCC TGCCAGGGCC TGGGCGATGG CATAGACCAC TGTGTAGGCG ATGTGTACTT TGCTTTCCTT CTGGAAAGGG TAGTGCTCAA CCCTCAGACT CTCATTTCCT GAGCAGAACT GGACTCCTGC AGCCGCTGTG CTTTTCCCTG CTGGCCCCCA GCTGCTATGG GGCTATGTGC ACTTGAAGGT GACCTCCCAG AATCTCCATA TGAACATGTC TTCTGGGGTC CTGGCAAGGC GCAGGGAAGT CGGGGGGCTG GCTGTTGCAA AGCAGGAGGC TGAATGAGCC CTGCAGAACC CAGAAGGCAC TGGCACAGTC AGGACCAGCG CCATGTGCAG AGTGTCCCAG CTGACACAGA CCTGGCCCAG GATGCCTCGG CCTGCCAGGC ACTGCAGGAT GAGTCTGAAG TGCTAGCTGT TAAGGAAAAC CAGGATGATG GTGGCCGTAC AGTTCTCCAT CAGCCAGATG ATGTTTTTCA TCTTCAGGGA CTCCCGGGTG GGGATGCAGA GTTGGAATTC AATGCAGACC CCAGCCTGGC CCAGCTCCTG GACGACCAGA GACCTGCCCT GCAGCTCAAA GTCGTCCCCC TGCGCCAGGA CGCTCACCCA AGACCATCGG AAGTGAAGCA TCAGCTGCAC CACTGCATGG GTGGTCGTGA GGTGACTGAG CAGGGTCTGA AAGAAGGATG GGAACTGGGT CCTGTCGCTG AGGCTGGCCA GTGTGGAGGA GGGGCTGACC TG	SEQ ID NO:17	P41180	Extracellular calcium-sensing receptor precursor	1078	164	460	264	25

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
190437 LG355	AC008754	Genomic clone	1068	GGAGTTTCCT CCTCTGAGTT TTCATCGTCT TTCTCTCCTG CCCAGACACC AGGAGCCTGA ATGGGGAACG ATTCTGTCAG CTACGAGTAT GGGGATTACA GCGACCTCTC GGACCGCCCT GTGGACTGCC TGGATGGCGC CTGCCTGGCC ATCGACCCGC TGCGCGTGGC CCCGCTCCCA CTGTATGCCG CCATCTTCCT GGTGGGGGTG CCGGGCAATG CCATGGTGGC CTGGGTGGCT GGGAAGGTGG CCCGCCGGAG GGTGGGTGCC ACCTGGTTGC TCCACCTGGC CGTGGCGGAT TTGCTGTGCT GTTTGTCTCT GCCCATCCTG GCAGTGCCCA TTGCCCGTGG AGGCCACTGG CCGTATGGTG CAGTGGGCTG TCGGGCGCTG CCCTCCATCA TCCTGCTGAC CATGTATGCC AGCGTCCTGC TCCTGGCAGC TCTCAGTGCC GACCTCTGCT TCCTGGCTCT CGGGCCTGCC TGGTGGTCTA CGGTTCAGCG GGCGTGCGGG GTGCAGGTGG CCTGTGGGGC AGCCTGGACA CTGGCCTTGC TGCTCACCGT GCCCTCCGCC ATCTACCGCC GGCTGCACCA GGAGCACTTC CCAGCCCGGC TGCAGTGTGT GGTGGACTAC GGCGGCTCCT CCAGCACCGA GAATGCGGTG ACTGCCATCC GGTTTCTTTT TGGCTTCCTG GGGCCCCTGG TGGCCGTGGC CAGCTGCCAC AGTGCCCTCC TGTGCTGGGC AGCCCGACGC TGCCGGCCGC TGGGCACAGC CATTGTGGTG GGGTTTTTTG TCTGCTGGGC ACCCTACCAC CTGCTGGGGC TGGTGCTCAC TGTGGCGGCC CCGAACTCCG CACTCCTGGC CAGGGCCCTG CGGGCTGAAC CCCTCATCGT GGGCCTTGCC CTCGCTCACA GCTGCCTCAA TCCCATGCTC TTCCTGTATT TTGGGAGGGC TCAACTCCGC CGGTCACTGC CAGCTGCCTG TCACTGGGCC CTGAGGGAGT CCCAGGGCCA GGACGAAAGT GTGGACAGCA AGAAATCCAC CAGCCATGAC CTGGTCTCGG AGATGGAG	GPCR C5L2 SEQ ID NO:33	P21730	C5A Anaphylatoxin chemotactic receptor	350	10	339	319	40

LS Cluster ID: Current (Original) LG NO.	Acc. No.	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
190438 LG6885	AL139287	Genomic clone	1086	CCAGCGGATC TTCAGGCGCT CTGTCCTGAG GCTGCCGTTG AACCTGCCCA CGTCGTGGAG CCTGGGCACT GAGCCCTGCC ACACCCACAG CTTCAGGTCG TACTCCATGT CCACGTTTCC GCTGCTGTCG AACCGCAGCG GCAGCCCGCC CACGTGGAAG GTCAGGTTGT ACATGTTCTC CAGGAGCTGC GGGCGGGGGA CGGGGCTGAG CCGCCAGCCA CCTCCAGCTC AGGCGTGGTG GCCGTGCCTG GTGGCCTGGG CACATGCAGA GGACAGCACA CCCCCATCTC CCGGGCTCAC CTGCCAGGGC TTCACGGGGT CCTGCGCGGG GCAGCCTGAG GCGTTGCACT GAAGAGTGTT GTGCAGGGCC TGGGCCACGC TATACACAGC TGCGTAGACA GAGAACGTCT GGTGGTGATT TAGCCCTGCG CTCACGTTCT GCAGCGTGAT GCAGTCACAC TGCGGGCAGC GCTGGCCCAC CACGTCCTCC TCCAGACCCT GCTCCCTCTC GCCCAGGGCA GAGCAGAAGG CCGGGTCGGT GGCCAGGGCC AGGTGCGTCT TCACGTACTG GGGGAACTCG TGCAGCTGGG CACCCCTCTG GAGGAAGCCA AGCACCGTGC CCATCTGGGC CATGCCGGGC AGCCCCATGA CCAGGTCAGA GGTCAGCCAG GCCTCGCTGG CCACCCACAC CTTGGGCGAG AGCCTGCTGC TGATGCTGTA GTTGAAGAGG GCGTGGGCGG CGTGCACGGA GGCGAACAGC AGCACCACCT GCACGCTGCT CTGGTTCACC TGGTGCAGGA CGTCCTGCAC CTTCCCCAGC CGCGAGTCAT CGGCACGGGG CAGCGGCACC AGGCCCTCGT GCGCGATGCA GATGCCGCGT GCCGCGGCCA GGGCCGAGAA GATGCTCAGG CCCTGCCGGC CGTACTCGTC GTCGCTGCCC AGGGCGGCCA CCCAGTTCCA GCCGAACTCC TGCAGCAGCT CCGCGGCGGC CGTCAGCTGC ACACGGTCGC TGGGCACGGT GCGGAAGAAG GAGGGGAAGG TCTCCCGGGC GCTCAGCAGC TCCATGCTAG CACCGTAGCT GACCTG		P02458	Procollagen alpha 1 (II) chain precursor	1418	360	710	350	31
190486 LG5968	AC018755	Genomic clone	377	GCCTGACAGG GCCATGGACA CCGGTCGTGG AGATGGCAAT GGCTCGGCTG TGCATTCAGC ATGTTCTGTA GATAGAATGA GTAAGACGTG GATGCGTTGG GTGGTGAGGA AAAGAGAAAT GAAAGGTGAG TTCTAGATCT GGACTTTAAC AACTGAATAT TTAATCTCCA AATAGATACA GAATATTGGA AGGTGGCAGG TCTGGGAGGC AGAGAACATC AACAGTTTGG TAAAACCATG TAGGTTTGGA GTCTATGGGA CTCCCACGGG GAGAGGTTTT ATCAGACTTG GTGCCAGGGA GAGGCCACGG CTGATAATTT AGATGAAAAG ACAGCATAAG GTTTAATCCC AGAGACTGGA TGATATCATT TATAGAA		M84562	Formyl peptide receptor-like receptor (FPRL1)	2631 bp	45	426	386	81

LS Cluster ID: Current (Original) LG NO.	Acc. No	Database Type	Sequence Length	Sequence	LS Cluster Name and Representative Sequence (SEQ ID NO)	Homolog Acc. No.	Homolog Name	Len.	From	To	Aligned	Percent
190774 (190488) LG263	AC007922	Genomic clone	540	CTTTTTTATA CAAAATATTT TCAAGAAAGC CTTTTGAAAG CGCTTGTGAC ACAATGGATA CAAAAGAGGA TTGACAAAGG AATTGAACCA CTGANAGCCA AAATGCAATT CTATACCAAA CTGATTTAGG ACCTGTTGCT GAGGAATAAA ATGAAAGGAC AATTGTGAAC AGAGNAATAT GGAGCCCAGC AAACAGCAAA AACCCCNTAA GAGAATGGCC AGTGACTTGG CTAATCTCCT GGCTCTAAGC AGTTCAACAT GTTCCCTTTG GTGAAGAGCT ACAGAATCTG ATTGGGAGAA GGAACCCATT TTGGAAGCAA TTGTATTGCT ATTCATCTTG GTTCTTGAGG AAAACATGAG ACTACTCTTT CTCCTCTGTC TCTCTGAATG AAAGGATGCA GGAACTTCTG TCGATGCAGA AAGAGATCTC CTTGAAGATA GTCTACCTCT GAATGAGTGT CCACAGATGT TGGAAGAGAC AGCAGTCAGT CCAGGATGGC TTTGGCACCT ACTGAGATGA TCACGCTTCC ACAGGCTCCA	Histamine H4 receptor (SEQ ID NO:37)	Q9Y5N1	Histamine H3 receptor	445	259	431	180	29
190557	AI806860	Dbest	574	TTTCCAAGAT TTATTTTTAT TATTAATTGA TATTAAATCT ATGACACTTC TGTTGTGTAT TTTTCAGAAT ATTGGTTCTT CTTTTATCTG AAGTGAGGAG CACCAGAGGA CTGATCACGG CATAGACATT GACTACAAGA CTCTGAACTT GCTGGTTGAT TGGGCCATAT GCCCACAACA TTGCTGAGGA GAAGGAGATG ATGAAATCCA CCCAACAGAG GACCACAGAG AAACTTACCA GCAGCAGTAT GGTCTGGATG GCCCGTTTCT CAGGGGAAGC TCCTGGGGAA GGACCATTGC TGTGAAGGTG GTGGGATCAC CTCTGAGGCC TGAATAAGAG AGTCACCATG TATGCAATTA AGAACAGCAG TATTCCTACC AGGAAAGCAT CCCTAAGTGT TGTCAGAATA AGAAACGTGG CCCTGAGGAT GAAGCTCATG GAGAAAACTG AGCAGTACTT ACCTATATTC AGTACATTTG TCTGGCTCAC ACTGGAAGCA GCTACAGTGT AGAAGATCCT GTTACTACTG AAAGACAAGT TGAGGAACCA TAAGAAGAAG AAACATGGAT AATG		Q62851	Pheromone receptor VN7, rat	273	99	269	175	36

WHAT IS CLAIMED IS:

1. An isolated polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence that is at least about 80% identical to the sequence set forth in Table 1.

2. The isolated polypeptide of claim 1, wherein the nucleotide sequence is set forth in Table 1.

3. An isolated nucleic acid molecule, or its complement, encoding the polypeptide of claim 1, wherein said nucleic acid molecule is operably linked to a heterologous promoter.

4. An expression vector comprising a nucleic acid molecule, or its complement, wherein the nucleic acid molecule encodes the polypeptide of claim 1.

5. A host cell comprising the expression vector of claim 4.

6. The host cell of claim 5, wherein the host cell is from a mammal.

7. A nucleic acid probe that specifically hybridizes with a nucleic acid molecule encoding the polypeptide of claim 1.

8. The nucleic acid probe of claim 7, wherein the nucleic acid is a DNA.

9. The nucleic acid probe of claim 7, wherein the nucleic acid is an RNA.

10. An expression vector comprising a nucleic acid molecule, or its complement, wherein the nucleic acid molecule selectively hybridizes to a sequence selected from Table 1, wherein the hybridization reaction is incubated overnight at 37°C in a solution comprising 40% formamide, 1 M NaCl and 1% SDS, and washed at 55°C in a solution comprising 0.5x SSC.

11. An antibody that selectively binds to the polypeptide of claim 1.

12. The antibody of claim 11, wherein said antibody is a monoclonal antibody.

13. The antibody of claim 11, wherein said antibody is a polyclonal antibody.

14. An antisense polynucleotide comprising a sequence capable of specifically hybridizing to a nucleic acid molecule encoding the polypeptide of claim 1.

15. A method for identifying a compound that modulates the expression of a polypeptide in a cell, wherein said polypeptide has at least 80% amino acid sequence identity to a polypeptide encoded by a nucleotide sequence selected from the group consisting of the sequences set forth in Table 1, the method comprising the steps of:

(a) culturing said cell in the presence of a modulator to form a first cell culture;

(b) contacting RNA or cDNA from the first cell culture with a probe which comprises a polynucleotide sequence encoding said polypeptide; and

(c) determining whether the amount of the probe which hybridizes to the RNA or cDNA from the first cell culture is increased or decreased relative to the amount of the probe which hybridizes to RNA or cDNA from a second cell culture grown in the absence of said modulator.

16. A method for identifying a compound that modulates the expression of at least two polypeptides in a cell, wherein each of said polypeptides has at least 80% amino acid sequence identity to a polypeptide encoded by a nucleotide sequence selected from the group consisting of the sequences set forth in Table 1, the method comprising the steps of:

(a) culturing said cell in the presence of a modulator to form a first cell culture;

(b) contacting RNA or cDNA from the first cell culture with at least two probes, each probe comprising a polynucleotide sequence encoding one of said polypeptides; and

(c) determining whether the amount of the probes which hybridizes to the RNA or cDNA from the first cell culture is increased or decreased relative to the amount of the probes which hybridizes to RNA or cDNA from a second cell culture grown in the absence of said modulator.

17. A method for identifying a compound that modulates the activity of a polypeptide, wherein said polypeptide has at least 80% amino acid sequence identity to a polypeptide encoded by a nucleotide sequence selected from the group consisting of the sequences set forth in Table 1, the method comprising the steps of:

(a) culturing cells expressing said polypeptide in the presence of a modulator to form a first cell culture; and

(b) measuring the activity of said polypeptide or second messenger activity in the first cell culture and determining whether the activity is increased or decreased relative to the activity of said polypeptide or second messenger activity from a second cell culture grown in the absence of said modulator.

18. A method for identifying a compound that modulates the activity of at least two polypeptides, wherein each of said polypeptides has at least 80% amino acid sequence identity to a polypeptide encoded by a nucleotide sequence selected from the group consisting of the sequences set forth in Table 1, the method comprising the steps of:

(a) culturing cells expressing said polypeptides in the presence of a modulator to form a first cell culture; and

(b) measuring the activity of said polypeptides or second messenger activity in the first cell culture and determining whether the activity is increased or decreased relative to the activity of said polypeptides or second messenger activity from a second cell culture grown in the absence of said modulator.

SEQUENCE LISTING

SEQ ID NO:1

189884

Cluster name: G protein-coupled receptor Ls189884 (putative GALR4 receptor)

SequenceID: LG610

Sequence: GGAGGGTACC TGCCCTCTGA TTCCCAGGAC TGGAGAACCA TCATCCCGGC TCTCTTGGTG GCTGTCTGCC TGGTGGGCTT CGTGGGAAAC CTGTGTGTGA TTGGCATCCT CCTTCACAAT GCTTGGAAAG GAAAGCCATC CATGATCCAC TCCCTGATTC TGAATCTCAG CCTGGCTGAT CTCTCCCTCC TGCTGTTTTC TGCACCTATC CGAGCTACGG CGTACTCCAA AAGTGTTTGG GATCTAGGCT GGTTTGTCTG CAAGTCCTCT GACTGGTTTA TCCACACATG CATGGCAGCC AAGAGCCTGA CAATCGTTGT GGTGGCCAAA GTATGCTTCA TGTATGCAAG TGGCCCAACC CAGCAAGTGG TTTTTCAACT ACCCCATTTG GTAATGGCGG TTGGCCTTTT GACTGGGGCT TACCTGTTA

SEQ ID NO:2

3098

Cluster name: Metabotropic glutamate receptor 6

SequenceID: NM_000843

Sequence: CGGAGGCCCG GGCAGGCCGG CTGAGCTAAC TCCCCAGAGC CAAAGTGGAA GGCGCGCCCC GAGCGCCTTC TCCCCAGGAC CCCGGTGTCC CTCCCCGCGC CCCGAGCCCG CGCTCTCCTT CCCCCGCCCT CAGAGCGCTC CCCGCCCCTC TGTCTCCCCG CAGCCCGCTA GACGAGCCGA TGGCGCGGCC CCGGAGAGCC CGGGAGCCGC TGCTCGTGGC GCTGCTGCCG CTGGCGTGGC TGGCGCAGGC GGGCCTGGCG CGCGCGGCGG GCTCTGTGCG CCTGGCGGGC GGCCTGACGC TGGGCGGCCT GTTCCCGGTG CACGCGCGGG GCGCGGCGGG CCGGGCGTGC GGGCCGCTGA AGAAGGAGCA GGGCGTGCAC CGGCTGGAGG CCATGCTGTA CGCGCTGGAC CGCGTCAACG CCGACCCCGA GCTGCTGCCC GGCGTGCGCC TGGGCGCGCG GCTGCTGGAC ACCTGCTCGC GGGACACCTA CGCGCTGGAG CAGGCGCTGA GCTTCGTGCA GGCGCTGATC CGCGGCCGCG GCGACGGCGA CGAGGTGGGC GTGCGCTGCC CGGGAGGCGT CCCTCCGCTG CGCCCCGCGC CCCCCGAGCG CGTCGTGGCC GTCGTGGGCG CCTCGGCCAG CTCCGTCTCC ATCATGGTCG CCAACGTGCT GCGCCTGTTT GCGATACCCC AGATCAGCTA TGCCTCCACA GCCCCGGAGC TCAGCGACTC CACACGCTAT GACTTCTTCT CCCGGGTGGT GCCACCCGAC TCCTACCAGG CGCAGGCCAT GGTGGACATC GTGAGGGCAC TGGGATGGAA CTATGTGTCC ACGCTGGCCT CCGAGGGCAA CTATGGCGAA AGTGGGGTTG AGGCCTTCGT TCAGATCTCC CGAGAGGCTG GGGGGGTCTG TATTGCCCAG TCTATCAAGA TTCCCAGGGA ACCAAAGCCA GGAGAGTTCA GCAAGGTGAT CAGGAGACTC ATGGAGACGC CCAACGCCCG GGGCATCATC ATCTTTGCCA ATGAGGATGA CATCAGGCGG GTCCTGGAGG CAGCTCGCCA GGCCAACCTG ACCGGCCACT TCCTGTGGGT CGGCTCAGAC AGCTGGGGAG CCAAGACCTC ACCCATCTTG AGCCTGGAGG ACGTGGCCGT TGGGGCCATC ACCATCCTGC CCAAAAGGGC CTCCATCGAC GGATTTGACC AGTACTTCAT GACTCGATCC CTGGAGAACA ACCGCAGGAA CATCTGGTTC GCCGAGTTCT GGGAAGAGAA TTTTAACTGC AAACTGACCA GCTCAGGTAC CCAGTCAGAT GATTCCACCC

GCAAATGCAC AGGCGAGGAA CGCATCGGCC GGGACTCCAC CTACGAGCAG GAGGGCAAGG TGCAGTTTGT GATTGATGCG GTGTATGCCA TTGCCCACGC CCTCCACAGC ATGCACCAGG CGCTCTGCCC TGGGCACACA GGCCTGTGCC CGGCGATGGA ACCCACCGAT GGGCGGATGC TTCTGCAGTA CATTCGAGCT GTCCGCTTCA ACGGCAGCGC AGGAACCCCT GTGATGTTCA ACGAGAACGG GGATGCGCCC GGGCGGTACG ACATCTTCCA GTACCAGGCG ACCAATGGCA GTGCCAGCAG TGGCGGGTAC CAGGCAGTGG GCCAGTGGGC AGAGACCCTC AGACTGGATG TGGAGGCCCT GCAGTGGTCT GGCGACCCCC ACGAGGTGCC CTCGTCTCTG TGCAGCCTGC CCTGCGGGCC GGGGGAGCGG AAGAAGATGG TGAAGGGCGT CCCCTGCTGT TGGCACTGCG AGGCCTGTGA CGGGTACCGC TTCCAGGTGG ACGAGTTCAC ATGCGAGGCC TGTCCTGGGG ACATGAGGCC CACGCCCAAC CACACGGGCT GCCGCCCCAC ACCTGTGGTG CGCCTGAGCT GGTCCTCCCC CTGGGCAGCC CCGCCGCTCC TCCTGGCCGT GCTGGGCATC GTGGCCACTA CCACGGTGGT GGCCACCTTC GTGCGGTACA ACAACACGCC CATCGTCCGG GCCTCGGGCC GAGAGCTCAG CTACGTCCTC CTCACCGGCA TCTTCCTCAT CTACGCCATC ACCTTCCTCA TGGTGGCTGA GCCTGGGGCC GCGGTCTGTG CCGCCCGCAG GCTCTTCCTG GGCCTGGGCA CGACCCTCAG CTACTCTGCC CTGCTCACCA AGACCAACCG TATCTACCGC ATCTTTGAGC AGGGCAAGCG CTCGGTCACA CCCCCTCCCT TCATCAGCCC CACCTCACAG CTGGTCATCA CCTTCAGCCT CACCTCCCTG CAGGTGGTGG GGATGATAGC ATGGCTGGGG GCCCGGCCCC CACACAGCGT GATTGACTAT GAGGAACAGC GGACAGTGGA CCCCGAGCAG GCCAGAGGGG TGCTCAAGTG CGACATGTCG GATCTGTCTC TCATCGGCTG CCTGGGCTAC AGCCTCCTGC TCATGGTCAC GTGCACAGTG TACGCCATCA AGGCCCGTGG CGTGCCCGAG ACCTTCAACG AGGCCAAGCC CATCGGCTTC ACCATGTACA CCACCTGCAT CATCTGGCTG GCATTCGTGC CCATCTTCTT TGGCACTGCC CAGTCAGCTG AAAAGATCTA CATCCAGACA ACCACGCTAA CCGTGTCCTT GAGCCTGAGT GCCTCGGTGT CCCTCGGCAT GCTCTACGTA CCCAAAACCT ACGTCATCCT CTTCCATCCA GAGCAGAATG TGCAGAAGCG AAAGCGGAGC CTCAAGGCCA CCTCCACGGT GGCAGCCCCA CCCAAGGGCG AGGATGCAGA GGCCCACAAG TAGCAGGGCA GGTGGGAACG GGACTGCTTG CTGCCTCTCC TTTCTTCCTC TTGCCTCGAG GTGGAAGCTG TATAGAGCCC GGGTCCACGG TGAACAGTCA GTGGCAGGGA GTTTGCCAAG ACCATGCTCC GCGTCGGTGG GGCTGGCCTT GAGAAGGAAC TGGACCCAGC TCTACCCCGA TTCCAGCATG TGAGCTTCAT GCTTCCTCAC CACAGACCAG ACTCGCTTCC CATGGTGGGA AACAGCCACC GAGAAGGTTC TAGCTCTAGA AAGGGACTAA ACTTATTCTC TCATCCGAAG TCCAAAGAGG ATGATGAAGC CCTGGGCTTT GCCTGGTTTG CGGGAGATTT CCTCCCCTCA GTCAACCCCC ATAACCTGGG GATTGGGCAG TGTGGAAGAA CGTGTAGACC CCAGAATGAA ACATGGGGTT GGAGTGGAGG AGGAGCTGTC TCAGCAAGAG GAGACCTGGG GCTGTGCATC TGGATGGAGG CACTCAGGCC TGGGTAGGAT TCCTCTGGCA CGGAGGGAGA GACCCTGGGT GAGACCCCTG TGAGCATGGG AAGGGCCTGC AGTGGGCGCG GGAGTGAGCT GAGGAACTGG GGTGCGCCCC CATGAGATTC CCAATGCCAT GGGCTTTCCC CCATCCCCCC GGGATTGGGC AAGGTCAGAC TTAGAGTACA GCTGTTTTCC TCCCCTCTGT GTACTCCCTT AAATCACCCC AACCTTGGCC AGGCATGGTG GCTCACACCT GTAATCCCAG CACTTTGGGA GGCCGAGGCA GGTGGATCAC CTGAGGTCCG GAGTTCGAGA CCAGCCTGGC CAATGTGGTG AAACCCTGTC TCTACTAAAA ATACAAAAAT TAGCCAGGTG TGATGGTGGG TGCCTGTAAT CCCAGTTACT TGGGAGGCTG

AGGCAGGAGA ATCGCTTGAA CCTGGGAGGT GGAGGTTGCA GTGAGCTGTG ATTGTGCCAC TGTACTCCAG CCTGGGTGAC AGAGCGAGAC TCTGTCTCAA AAAAACAAAA CAAAAAAACA CCAAAAAAAC CCCCAAACCT GAAGAAATTC AGATACACGT GTGTAATGTT AGTGATGTGA GAACAAGGAG CAGGGGTGCA TTTGTGTTGT GTTCGGGTTG GGGATGGGTT TAGGAGCTCC AGGTTGGGAG CAGTGACAGA GAGTCATGGC CGTGGTGAGG GTGAATCCCA AGTGGATGGC TCAGGACGGG TATGGAAACC CTTCATTCCT CATAGGTACT GGGAAGTCCA TTTGCAAGCT GAGCGCCAGG CCTGGGGAGG AAGAGGCTTG GGCTGCAGAT GCACGCACAT TTGTTTTTCA CTGATAGTTT TTACAAAAAG CTTGGTTTAA GTTATGGAAT TTTATGTCCC TGGGAGTAGA ATTTACATTT GTTAAATTGA CCACTGTTTA AGATCAGTAT ACATTCTCTA GTCTGTGATG TCTGGAGCTA GTTTTGAGGG TGAACCACAC TTTATCCAAC ATACAAACTT TCCCATGCAG CTTCTCTGGT GCGCAGTTGG TTTTGACCGT GGGACTAGGT GCTTCTGCAG GTTTTAAGTA ATTAACTTAA AAGCTTCTCC TCTGAGAAAC ATTTCTGTTG CGCTACTGAC TCTCCTTCTC CACATTTGTT GTGTTCCTAG GGCTTCTCTA TAGTGCACAT TAGGACGTTT CATTTGTTGC TGAATGCTTT CCAGAATTAT TTATTCCATA GGGTTTCTCT CCTGTGCAGC TCTCTCATGG GTAATGGGGC GTGTTTTCTT GCCAAAGGCG GTTCCACCCT CGTGATTGTA TAGGGCTCTT CTCCTGTATG AACTCTGAGA TCAGTGAGCT CTGATCTCCA AGGGAAAGTT TTCCTGCATT TGCTGTTTTC TCATGTCTCT CCCAGTGTGA ATTCTCTGGC TTCTAGCTGA AAACTTTTCC ACAGTTTTAC ATTCATGTGG TTTTCTCCAC TGTGAACTCT GTGATTCAGA ATCAGAAGCA GTTCTTAGTA GAGGCATTTC TACACTGATT GCACTGAGGA TATCTCCCCA GTGTGAAGTT TCTGGCATAG AGTCCTGGCT TCCCGCAGAC GACTTTCACA CTCTGCCATG TTCATGCCTG TGGGCCTCTC TGGCAGGAAC TCTGATGCAC CGCGAGGCCC ATGTACTCCT GTGGCTTTCT CACATTCGGT CTACTTGCAG GGTATCTCCA CAGCATGCAC CATTCTGGGT ACAGGGGGAC ATCCTCTGTT ACTGAAGATG TTGTCATATT TAGTACCTTC ACAAGGTTTC TCTCCTTCCA GAATTTTCTG ATGTACACAA ATAACTGACT TCCACAAGAG GGCTTTTCCA CACTCGGTGT GTGCATACAG TTTCTGCCTG TGATCATTTC TTTATGTTAT TATTTTATTT TTTCGAGATA GGGTCTTGCT CAATTTCTTA GGCTGGAGTG CAGTGGCACG ATCATAGCTC ACTGAAGTTT CGACCTGGGC TCAAGCAATC CTCCCGCTTC AGCCTCCTGA GTAGCTGGTG CGCACGACCA TACCCAGCTA ATGTTTTATT TTTTGTAGAG ACGAGGTCTC ACTATGTTGC CCAGGCTGGT CTCGAACTTC TGAGCTCGAG CGATCCTCCT GCCTCCACCT CCCAAAGTGT TCGGATTACA AACGTGAGCC ATCGCACCTA GCCTCTTTGA TCATTTCTGT GGTGTTCAGT GGGGGTTGAC AGCTCCCTAA AGATTTTCCT GTTTTTTTGC ATGCATGGGT

SEQ ID NO:3

22315

Cluster name: G protein-coupled receptor GPR92

SequenceID: NM_020400

Sequence: ATGTTAGCCA ACAGCTCCTC AACCAACAGT TCTGTTCTCC CGTGTCCTGA CTACCGACCT ACCCACCGCC TGCACTTGGT GGTCTACAGC TTGGTGCTGG CTGCCGGGCT CCCCCTCAAC GCGCTAGCCC TCTGGGTCTT CCTGCGCGCG CTGCGCGTGC ACTCGGTGGT GAGCGTGTAC ATGTGTAACC TGGCGGCCAG CGACCTGCTC TTCACCCTCT CGCTGCCCGT TCGTCTCTCC TACTACGCAC TGCACCACTG GCCCTTCCCC GACCTCCTGT

GCCAGACGAC GGGCGCCATC TTCCAGATGA ACATGTACGG
CAGCTGCATC TTCCTGATGC TCATCAACGT GGACCGCTAC
GCCGCCATCG TGCACCCGCT GCGACTGCGC CACCTGCGGC
GGCCCCGCGT GGCGCGGCTG CTCTGCCTGG GCGTGTGGGC
GCTCATCCTG GTGTTTGCCG TGCCCGCCGC CCGCGTGCAC
AGGCCCTCGC GTTGCCGCTA CCGGGACCTC GAGGTGCGCC
TATGCTTCGA GAGCTTCAGC GACGAGCTGT GGAAAGGCAG
GCTGCTGCCC CTCGTGCTGC TGGCCGAGGC GCTGGGCTTC
CTGCTGCCCC TGGCGGCGGT GGTCTACTCG TCGGGCCGAG
TCTTCTGGAC GCTGGCGCGC CCCGACGCCA CGCAGAGCCA
GCGGCGGCGG AAGACCGTGC GCCTCCTGCT GGCTAACCTC
GTCATCTTCC TGCTGTGCTT CGTGCCCTAC AACAGCACGC
TGGCGGTCTA CGGGCTGCTG CGGAGCAAGC TGGTGGCGGC
CAGCGTGCCT GCCCGCGATC GCGTGCGCGG GGTGCTGATG
GTGATGGTGC TGCTGGCCGG CGCCAACTGC GTGCTGGACC
CGCTGGTGTA CTACTTTAGC GCCGAGGGCT TCCGCAACAC
CCTGCGCGGC CTGGGCACTC CGCACCGGGC CAGGACCTCG
GCCACCAACG GGACGCGGGC GGCGCTCGCG CAATCCGAAA
GGTCCGCCGT CACCACCGAC GCCACCAGGC CGGATGCCGC
CAGTCAGGGG CTGCTCCGAC CCTCCGACTC CCACTCTCTG
TCTTCCTTCA CACAGTGTCC CCAGGATTCC GCCCTCTGA

SEQ ID NO:4

30875

Cluster name: G protein-coupled receptor GPR87

SequenceID: NM_023915

Sequence: GGCACGAGGG TTTCGTTTTC ATGCTTTACC AGAAAATCCA
CTTCCCTGCC GACCTTAGTT TCAAAGCTTA TTCTTAATTA
GAGACAAGAA ACCTGTTTCA ACTTGAAGAC ACCGTATGAG
GTGAATGGAC AGCCAGCCAC CACAATGAAA GAAATCAAAC
CAGGAATAAC CTATGCTGAA CCCACGCCTC AATCGTCCCC
AAGTGTTTCC TGACACGCAT CTTTGCTTAC AGTGCATCAC
AACTGAAGAA TGGGGTTCAA CTTGACGCTT GCAAAATTAC
CAAATAACGA GCTGCACGGC CAAGAGAGTC ACAATTCAGG
CAACAGGAGC GACGGGCCAG GAAAGAACAC CACCCTTCAC
AATGAATTTG ACACAATTGT CTTGCCGGTG CTTTATCTCA
TTATATTTGT GGCAAGCATC TTGCTGAATG GTTTAGCAGT
GTGGATCTTC TTCCACATTA GGAATAAAAC CAGCTTCATA
TTCTATCTCA AAAACATAGT GGTTGCAGAC CTCATAATGA
CGCTGACATT TCCATTTCGA ATAGTCCATG ATGCAGGATT
TGGACCTTGG TACTTCAAGT TTATTCTCTG CAGATACACT
TCAGTTTTGT TTTATGCAAA CATGTATACT TCCATCGTGT
TCCTTGGGCT GATAAGCATT GATCGCTATC TGAAGGTGGT
CAAGCCATTT GGGGACTCTC GGATGTACAG CATAACCTTC
ACGAAGGTTT TATCTGTTTG TGTTTGGGTG ATCATGGCTG
TTTTGTCTTT GCCAAACATC ATCCTGACAA ATGGTCAGCC
AACAGAGGAC AATATCCATG ACTGCTCAAA ACTTAAAAGT
CCTTTGGGGG TCAAATGGCA TACGGCAGTC ACCTATGTGA
ACAGCTGCTT GTTTGTGGCC GTGCTGGTGA TTCTGATCGG
ATGTTACATA GCCATATCCA GGTACATCCA CAAATCCAGC
AGGCAATTCA TAAGTCAGTC AAGCCGAAAG CGAAAACATA
ACCAGAGCAT CAGGGTTGTT GTGGCTGTGT TTTTTACCTG
CTTTCTACCA TATCACTTGT GCAGAATTCC TTTTACTTTT
AGTCACTTAG ACAGGCTTTT AGATGAATCT GCACAAAAAA
TCCTATATTA CTGCAAAGAA ATTACACTTT TCTTGTCTGC
GTGTAATGTT TGCCTGGATC CAATAATTTA CTTTTTCATG
TGTAGGTCAT TTTCAAGAAG GCTGTTCAAA AAATCAAATA
TCAGAACCAG GAGTGAAAGC ATCAGATCAC TGCAAAGTGT

GAGAAGATCG GAAGTTCGCA TATATTATGA TTACACTGAT GTGTAGGCCT TTTATTGTTT GTTGGAATCG ATATGTACAA AGTGTAAATA AATGTTTCTT TTCATTATCC TTAAAAAAAA AA

SEQ ID NO:5

54602

Cluster name: Pheromone receptor (PHRET) pseudogene

SequenceID AF253316

Sequence: TCTGACAGAC AACACCTTTT TGCTTTTCTT CCACATCTTC ACACTCCTTC AGGATCAAAA ACCTAAGCCA CATGACTGGA TGAGCCGTCA CTTGGCCTTC ATTCGGGTAG TGATGGTCCT CACTGTAGTG GATGTTTTGC CTCCAGATAT GCTTGAATCA CTGCATTTTG GGAATAACTT CAAATGCAAG TCCTTGATCT AAATAAACAG AATGACGAAG GGCCTATGTT TCTATACCAC CTGTCTCCTG AATATACACC AGGCCAGCAT AATCAGCCTC AGCAACTTCT GGTTGGAAAG CTTTAAACAT AAATTTACAA ATAACATTGT CAGTGTCCTC TTTTTTCTTT TTTGTTCCCT CAATTTGTCT TTCAGTAGTG ACATAATATT CTTCACTGTG GCTTCTTCCA TTGTGACCCA GACCAATCTA CTTAAGGTCC GCAAATACTG CTCACGTTCT CCCATGAAAT CCATCATGTG GGGAGTGTTT TCCTTGTAGG ATTACGCTGC TCTCAAGTGC ATACATGATG ATCTTTTTGT CCAAGCATCA GAAGTGATCC CAGCATCTTC ACAGTACCAG CCTTTCCCCA AGATCCTCGC CAGAGAAAAG GGTTACCCAG ATCATCCTGC CACTGGTGAA TTGCTTTGTT GTCATGTTCT GGGTGGACCT TATCATCTCA TCCTCTTCAT CCCTGTTATG GACGTATAAC CCAGTCATCC TGAGCATCTA GAACCTTGTT GCCTGTGTCT ATGCCACTCT CGTTCCATTG GTACAAATCC GCTCTGATAA AAGAATAGTC AATATTCTCC AAAAAATGGA ATTAAAGTGC TATAATTTTT TAATGTGTTG GTGATGAAAA ATATTTCTAA AAATTAGTCT CATTCTATAG TTAAATTGTT CAAGTAGCCC CAGATTTAGC TTACTGAGTT TAAATAAAAT GCGTGGAATT ACACTTTTAT TATATTTTTA TGCTTCTGAA ACTGAGGCAT CTAAGGACTA TGTAGTTTCT TCAGTTCAAT GTTCACCATA GATTGACATT TCAGATATCA AGTCTTTTGC ACTTTTATTT TTATGTTAAC TTTGTACAAG AAAATGTTTC TCTCTTTTTG AAGTACATTC TTAAAAAATT TGTTTTGGTA TCAATCTCTC AATGTTTTTA CTTTTGAAAA TATTTACTTA CTCTGTTTAT GAATGATACT TTAGCTCAAT ATTCAATTCT AGCTTTTAAG CCATGCTTGC TCATTGTACC TCCCTGACTA AAAAAAATTA TGTCTATTTG GATTTTAAAT TTAATCTAGA ATTCATTTTA ACG

SEQ ID NO:6

55728

Cluster name: ETL protein

SequenceID: NM_022159

Sequence: GTGAAATTTA AACTCCAGTC CTGTGGCGAA AATGCTAATT GCACTAACAC AGAAGGAAGT TATTATTGTA TGTGTGTACC TGGCTTCAGA TCCAGCAGTA ACCAAGACAG GTTTATCACT AATGATGGAA CCGTCTGTAT AGAAAATGTG AATGCAAACT GCCATTTAGA TAATGTCTGT ATAGCTGCAA ATATTAATAA AACTTTAACA AAAATCAGAT CCATAAAAGA ACCTGTGGCT TTGCTACAAG AAGTCTATAG AAATTCTGTG ACAGATCTTT CACCAACAGA TATAATTACA TATATAGAAA TATTAGCTGA ATCATCTTCA TTACTAGGTT ACAAGAACAA CACTATCTCA

GCCAAGGACA CCCTTTCTAA CTCAACTCTT ACTGAATTTG
TAAAAACCGT GAATAATTTT GTTCAAAGGG ATACATTTGT
AGTTTGGGAC AAGTTATCTG TGAATCATAG GAGAACACAT
CTTACAAAAC TCATGCACAC TGTTGAACAA GCTACTTTAA
GGATATCCCA GAGCTTCCAA AAGACCACAG AGTTTGATAC
AAATTCAACG GATATAGCTC TCAAAGTTTT CTTTTTTGAT
TCATATAACA TGAAACATAT TCATCCTCAT ATGAATATGG
ATGGAGACTA CATAAATATA TTTCCAAAGA GAAAAGCTGC
ATATGATTCA AATGGCAATG TTGCAGTTGC ATTTTTATAT
TATAAGAGTA TTGGTCCTTT GCTTTCATCA TCTGACAACT
TCTTATTGAA ACCTCAAAAT TATGATAATT CTGAAGAGGA
GGAAAGAGTC ATATCTTCAG TAATTTCAGT CTCAATGAGC
TCAAACCCAC CCACATTATA TGAACTTGAA AAAATAACAT
TTACATTAAG TCATCGAAAG GTCACAGATA GGTATAGGAG
TCTATGTGCA TTTTGGAATT ACTCACCTGA TACCATGAAT
GGCAGCTGGT CTTCAGAGGG CTGTGAGCTG ACATACTCAA
ATGAGACCCA CACCTCATGC CGCTGTAATC ACCTGACACA
TTTTGCAATT TTGATGTCCT CTGGTCCTTC CATTGGTATT
AAAGATTATA ATATTCTTAC AAGGATCACT CAACTAGGAA
TAATTATTTC ACTGATTTGT CTTGCCATAT GCATTTTTAC
CTTCTGGTTC TTCAGTGAAA TTCAAAGCAC CAGGACAACA
ATTCACAAAA ATCTTTGCTG TAGCCTATTT CTTGCTGAAC
TTGTTTTTCT TGTTGGGATC AATACAAATA CTAATAAGCT
CTTCTGTTCA ATCATTGCCG GACTGCTACA CTACTTCTTT
TTAGCTGCTT TTGCATGGAT GTGCATTGAA GGCATACATC
TCTATCTCAT TGTTGTGGGT GTCATCTACA ACAAGGGATT
TTTGCACAAG AATTTTTATA TCTTTGGCTA TCTAAGCCCA
GCCGTGGTAG TTGGATTTTC GGCAGCACTA GGATACAGAT
ATTATGGCAC AACCAAAGTA TGTTGGCTTA GCACCGAAAA
CAACTTTATT TGGAGTTTTA TAGGACCAGC ATGCCTAATC
ATTCTTGTTA ATCTCTTGGC TTTTGGAGTC ATCATATACA
AAGTTTTTCG TCACACTGCA GGGTTGAAAC CAGAAGTTAG
TTGCTTTGAG AACATAAGGT CTTGTGCAAG AGGAGCCCTC
GCTCTTCTGT TCCTTCTCGG CACCACCTGG ATCTTTGGGG
TTCTCCATGT TGTGCACGCA TCAGTGGTTA CAGCTTACCT
CTTCACAGTC AGCAATGCTT TCCAGGGGAT GTTCATTTTT
TTATTCCTGT GTGTTTTATC TAGAAAGATT CAAGAAGAAT
ATTACAGATT GTTCAAAAAT GTCCCCTGTT GTTTTGGATG
TTTAAGGTAA ACATAGAGAA TGGTGGATAA TTACAACTGC
ACAAAAATAA AAATTCCAAG CTGTGGATGA CCAATGTATA
AAAATGACTC ATCAAATTAT CCAATTATTA ACTACTAGAC
AAAAAGTATT TTAAATCAGT TTTTCTGTTT ATGCTATAGG
AACTGTAGAT AATAAGGTAA AATTATGTAT CATATAGATA
TACTATGTTT TTCTATGTGA AATAGTTCTG TCAAAAATAG
TATTGCAGAT ATTTGGAAAG TAATTGGTTT CTCAGGAGTG
ATATCACTGC ACCCAAGGAA AGATTTTCTT TCTAACACGA
GAAGTATATG AATGTCCTGA AGGAAACCAC TGGCTTGATA
TTTCTGTGAC TCGTGTTGCC TTTGAAACTA GTCCCCTACC
ACCTCGGTAA TGAGCTCCAT TACAGAAAGT GGAACATAAG
AGAATGAAGG GGCAGAATAT CAAACAGTGA AAAGGGAATG
ATAAGATGTA TTTTGAATGA ACTGTTTTTT CTGTAGACTA
GCTGAGAAAT TGTTGACATA AAATAAAGAA TTGAAGAAAC

SEQ ID NO:7

160221

Cluster name: G Protein-Coupled Receptor GPR27

SequenceID: NM_018971

Sequence: ATGGCGAACG CGAGCGAGCC GGGTGGCAGC GGCGGCGGCG

AGGCGGCCGC CCTGGGCCTC AAGCTGGCCA CGCTCAGCCT GCTGCTGTGC GTGAGCCTAG CGGGCAACGT GCTGTTCGCG CTGCTGATCG TGCGGGAGCG CAGCCTGCAC CGCGCCCCGT ACTACCTGCT GCTCGACCTG TGCCTGGCCG ACGGGCTGCG CGCGCTCGCC TGCCTCCCGG CCGTCATGCT GGCGGCGCGG CGTGCGGCGG CCGCGGCGGG GGCGCCGCCG GGCGCGCTGG GCTGCAAGCT GCTCGCCTTC CTGGCCGCGC TCTTCTGCTT CCACGCCGCC TTCCTGCTGC TGGGCGTGGG CGTCACCCGC TACCTGGCCA TCGCGCACCA CCGCTTCTAT GCAGAGCGCC TGGCCGGCTG GCCGTGCGCC GCCATGCTGG TGTGCGCCGC CTGGGCGCTG GCGCTGGCCG CGGCCTTCCC GCCAGTGCTG GACGGCGGTG GCGACGACGA GGACGCGCCG TGCGCCCTGG AGCAGCGGCC CGACGGCGCC CCCGGCGCGC TGGGCTTCCT GCTGCTGCTG GCCGTGGTGG TGGGCGCCAC GCACCTCGTC TACCTCCGCC TGCTCTTCTT CATCCACGAC CGCCGCAAGA TGCGGCCCGC GCGCCTGGTG CCCGCCGTCA GCCACGACTG GACCTTCCAC GGCCCGGGCG CCACCGGCCA GGCGGCCGCC AACTGGACGG CGGGCTTCGG CCGCGGGCCC ACGCCGCCCG CGCTTGTGGG CATCCGGCCC GCAGGGCCGG GCCGCGGCGC GCGCCGCCTC CTCGTGCTGG AAGAATTCAA GACGGAGAAG AGGCTGTGCA AGATGTTCTA CGCCGTCACG CTGCTCTTCC TGCTCCTCTG GGGGCCCTAC GTCGTGGCCA GCTACCTGCG GGTCCTGGTG CGGCCCGGCG CCGTCCCCCA GGCCTACCTG ACGGCCTCCG TGTGGCTGAC CTTCGCGCAG GCCGGCATCA ACCCCGTCGT GTGCTTCCTC TTCAACAGGG AGCTGAGGGA CTGCTTCAGG GCCCAGTTCC CCTGCTGCCA GAGCCCCCGG ACCACCCAGG CGACCCATCC CTGCGACCTG AAAGGCATTG GTTTATGA

SEQ ID NO:8

160314

Cluster name: G protein-coupled receptor Ls160314

SequenceID: ENSMDNA221753

Sequence: ATGAAGATCA AATATGACTT CCTATATGAA AAGGAACACA TCTGCTGCTT AGAAGAGTGG ACCAGCCCTG TGCACCAGAA GATCTACACC ACCTTCATCC TTGTCATCCT CTTCCTCCTG CCTCTTATGG TGATGCTTAT TCTGTACAGT AAAATTGGTT ATGAACTTTG GATAAAGAAA AGAGTTGGGG ATGGTTCAGT GCTTCGAACT ATTCATGGAA AAGAAATGTC CAAAATAGCC AGGAAGAAGA AACGAGCTGT CATTATGATG GTGACAGTGG TGGCTCTCTT TGCTGTGTGC TGGGCACCAT TCCATGTTGT CCATATGATG ATTGAATACA GTAATTTTGA AAAGGAATAT GATGATGTCA CAATCAAGAT GATTTTTGCT ATCGTGCAAA TTATTGGATT TTCCAACTCC ATCTGTAATC CCATTGTCTA TGCATTTATG AATGAAAACT TCAAAAAAAA TGTTTTGTCT GCAGTTTGTT ATTGCATAGT AAATAAAACC TTCTCTCCAG CACAAAGGCA TGGAAATTCA GGAATTACAA TGATGCGGAA GAAAGCAAAG TTTTCCCTCA GAGAGAATCC AGTGGAGGAA ACCAAAGGAG AAGCATTCAG TGATGGCAAC ATTGAAGTCA AATTGTGTGA ACAGACAGAG GAGAAGAAAA AGCTCAAACG ACATCTTGCT CTCTTTAGGT CTGAACTGGC TGAGAATTCT CCTTTAGACA GTGGGCATTA A

SEQ ID NO:9

160324

Cluster name: G protein-coupled receptor GPR86

SequenceID: NM_023914

Sequence: AACAGTATTT TCCTTTTCAA CACATCTATT GAAAGTGTTG GATAAATGCA GGATGTTAAT ATGCTATAAA CATAAAGTCT GTTTTTAAAA AATAGCATTT GAAAATCATG AAGGGCTTTT TGTTTTCTTT TGTTTGTATA TATGTTTATT GGTAACAGGT GACACTGGAA GCAATGAACA CCACAGTGAT GCAAGGCTTC AACAGATCTG AGCGGTGCCC CAGAGACACT CGGATAGTAC AGCTGGTATT CCCAGCCCTC TACACAGTGG TTTTCTTGAC CGGCATCCTG CTGAATACTT TGGCTCTGTG GGTGTTTGTT CACATCCCCA GCTCCTCCAC CTTCATCATC TACCTCAAAA ACACTTTGGT GGCCGACTTG ATAATGACAC TCATGCTTCC TTTCAAAATC CTCTCTGACT CACACCTGGC ACCCTGGCAG CTCAGAGCTT TTGTGTGTCG TTTTTCTTCG GTGATATTTT ATGAGACCAT GTATGTGGGC ATCGTGCTGT TAGGGCTCAT AGCCTTTGAC AGATTCCTCA AGATCATCAG ACCTTTGAGA AATATTTTTC TAAAAAAACC TGTTTTTGCA AAAACGGTCT CAATCTTCAT CTGGTTCTTT TTGTTCTTCA TCTCCCTGCC AAATATGATC TTGAGCAACA AGGAAGCAAC ACCATCGTCT GTGAAAAAGT GTGCTTCCTT AAAGGGGCCT CTGGGGCTGA AATGGCATCA AATGGTAAAT AACATATGCC AGTTTATTTT CTGGACTGTT TTTATCCTAA TGCTTGTGTT TTATGTGGTT ATTGCAAAAA AAGTATATGA TTCTTATAGA AAGTCCAAAA GTAAGGACAG AAAAAACAAC AAAAAGCTGG AAGGCAAAGT ATTTGTTGTC GTGGCTGTCT TCTTTGTGTG TTTTGCTCCA TTTCATTTTG CCAGAGTTCC ATATACTCAC AGTCAAACCA ACAATAAGAC TGACTGTAGA CTGCAAAATC AACTGTTTAT TGCTAAAGAA ACAACTCTCT TTTTGGCAGC AACTAACATT TGTATGGATC CCTTAATATA CATATTCTTA TGTAAAAAAT TCACAGAAAA GCTACCATGT ATGCAAGGGA GAAAGACCAC AGCATCAAGC CAAGAAAATC ATAGCAGTCA GACAGACAAC ATAACCTTAG GCTGACAACT GTACATAGGG TTAACTTCTA TTTATTGATG AGACTTCCGT AGATAATGTG GAAATCAAAT TTAACCAAGA AAAAAAGATT GGAACAAATG CTCTCTTACA TTTTATTATC CTGGTGTACA GAAAAGATTA TATAAAATTT AAATCCACAT AGATCTATTC ATAAGCTGAA TGAACCATTA CTAAGAGAAT GCAACAGGAT ACAAATGGCC ACTAGAGGTC ATTATTTCTT TCTTTCTTTT TTTTTTTTTT AATTTCAAGA GCATTTCACT TTAACATTTT GGAAAAGACT AAGGAGAAAC GTATATCCCT ACAAACCTCC CCTCCAAACA CCTTCTCACA TTCTTTTCCA CAATTCACAT AACACTACTG CTTTTGTGCC CCTTAAATGT AGATATGTGC TGAAAGAAAA AAAAAACGCC CAACTCTTGA AGTCCATTGC TGAAAACTGC AGCCAGGGGT TGAAAGGGAT GCAGACTTGA AGAGTCTGAG GAACTGAAGT GGGTCAGCAA GACCTCTGAA ATCCTGGGTA AAGGATTTTC TCCTTACAAT TACAAACAGC CTCTTTCACA TTACAATAAT ATACCATAGG AGGCACAAGC ACCATTATTA AGCCACTTTG CTTACACCTT AAGTGTGTAC AATTCAAGTG TGAGAATGCT GTGTTAACTA TTCTTTGGAA TTCTCCTTCT GTCCAGCAAA TACTCTAATG ATGGTTAAAC ATGGCACCTA CTCAGCAATG CCTTCCTGGA CCACAACCCC TATCCCCCTG CCCCACCCTC CTCATTAAAA ACAAATACTT CTACTGTTTG GGTGTGTGAT AGGGTTCTCA ATGCAGATCT CCCTTTTCTA GTTAGCTATA TTCTTGACTG CATCCGCTAA AAATGTTAAA GCTTCTTGAG AGACAGACAT GCCAGATTTT CTTGGTATCT CCCATAATAC GACCTACAGT CCATGGTCTA CAGATGTTTT AAATAGAATT GCTATTCTCG ATACATACAA AGACGTAATT GCTGACCCAC AATCAGTAAC ATCCATATTG GGAGATTTTT CAAAGGATGG TGACCCTGCT TGTATTTATT TACCTTGGTA TTTTTTCTTG CATCCTTCTG TGATTCAAAA AAGTAAAATG TGGCTTTCTG AAATGATGGA TAAGAGTCTA CATCTTCTAG AAAAAATACA

TAAAGGAGTA GTTAAGCTCT GTAAATGTGC CACGAGCTCC AACACGACCA TCGTAGGGTG AAGCCCACGT TTTCTTCCAT GGCCTCAAAG GCCCTAGAAC TTGCCTACCT TTCTGGCCTT ACCTCCTAGC TACTTATCCA TCTCTTGAAC TTTATACTCT TGTATAAATT TCTAACTTTC AGAAAATGCC ATACTCTGTT TTGGCACCAC ACATGTATAT TTCCCCCTGG TACACTTGGA AGACTCTTAT CCATCTGTGA AACCCTATGT TGTCATCACT TGGTCCATGA AATATTACCT GGCCAATATC CCACCATCAC CTCAAACCCA ATCACCCCCT CCTCTGTATG CTGTCACACC TATATTATTA AACTTATCAC ATTGCATTGT AATTACTTCC

SEQ ID NO:10

160458

Cluster name: G protein-coupled receptor Ls160458

SequenceID: AI733823

Sequence: TTTAAATTTA AAAACTTTAT TGGAATAGCA TGTTAGCAGC AGTGAACAGG GCATGGCACA GAAGGTTTCC AAAACAAGTT TAGCATGAAG GATGCCATAT GCTGTTGCCA ACAACTAGAA CACGGTGACT AAAGACACAG TTCTGAATGT CCAGCACAAC CTCTGGCCTG CAACTATGTT CAGTGATGAT GATAAACAAG GTGGTGACTT GGAAGGAATC CCTATGTCAA GTGAGAAAAA AAAATGATGT CTGACCTCCT TATATATGTA AAAAATATAC CTTCAGAGTC CGTCAGTAAG CTGGAAGAAG TGGATGTTGA AGTTTTTAAC ATCGATGATG GGTCTCCAGT TGTTCATCAA CCCATGGTGA AATAGCTGAA CGGTTCTGAA TCAAAGGTGA TCCTAATAGT GAAGACATTA ACATTGCAGA AAAAGTGCCT ACAGATTATA TGGTGAAAAT ACGTGATGGG CTTCTTGAAG GACTAGAGCA GTGTGTATTC AAAACAGAAC AAGAAATCAC GTCAGTTTAT

SEQ ID NO:11

160833

Cluster name: 5-HT5B receptor

SequenceID: AJ308679

Sequence: CCCCCTCCAC GCCCGCACCT GCCCGGTCCA CGCCGAACTC ACTGAGGACT CGTGTGCCCC CTGCCCTGGA GCTGCGATCC CAAGCGCCGT GGAGGCCGCT AGCCTTTCAG TGGCCACCGC CGGCGTTGCC CTTGCCCTGG GACCCGAGAC CAGCAGCAGG ACCCGGGACC CCAAGCCCGA GAGGGATACT CGGTTCGACC CCGAGCGGCG CCGTCCTGCC GGGCCGAGGG CCGCCCTTCT CTGTCTTCAC GGTCCTGGTG GTGACGCTGC TAGTGCTGCT GATCGCCGCC ACTTTCCTGT GGAACCTGCT GGTTCCGGTC ACCATCCCGC GGGTCCGTGC CTTCCACCCG GTGCCGCATA ACTTGGTGGC CTCGACGGCC GTCTCGGACG AACTAGTGGC AGCGCTGGCG ATGCCACCGA GCCTGGCGAG TGAGCTGTCG ACCGGGCGAC GTCGGCTGCT GGGCCGGAGC CTGTGCCACG TGTGGATCTC CTTCGACGCC GGAGCCTGTG CCACGTGTGG ATCTCCTTCC ACGGCTGTGC TGCCCCGCCG GCCTCGGGAA CGTGGCGGCC ATCGCCCTGG GCCGCGACGG GGCCATCACA CGGCACCTGC AGCACACGCT GCGCACCTGC AGCCGCGCCT CGTTGCTCAT GATCGCGCTC ACCCGGGTGC CGTCGGCGCT CATCGCCCTC GCGCCGCTGC TCTTTGGCCG GGGCGAGGTG TGCGACGCTC GGCTCCAGCG CTGCCAGGTG AGCCGGGAAC CCTCCTATGC CGCCTTCTCC ACCCGCGGCG CCTTCCACCT GCCGCTTGGC GTGGTGCCGT TTGTCTACCG GAAGATCTAC

GAGGCGGCCA AGTTTCGTTT CGGCCGACGC CGGAGAGCTG TGCTGCCGTT GCCGGCCACC ATGCAGGTGA GGGGTGGGCT GAGGAACGTT GCTTTGGCGA AGCGGTTGCT AGAGAAGGAG GCGGCTTCGC GAATGGC

SEQ ID NO:12

162615

Cluster name: G protein-coupled receptor Ls162615

SequenceID: BF115152

Sequence: TTGAAGCCAC TGAGACATTC TTGTTTTATT CCCAGACCCC TAAATCAGAA AACCCGATCG AATACTGAGC ATAATTTCTT CATTGACATT TGTCTCTAAA TGTCAAGTTG TTCTGGAAAT TTTTTCTTGA TTTTTNGATT CATTGCCTTA TTCATTTGAG ACAAACTGAG TTAGCATGAT GTTGTCGGAG GAATCTCCAG TATGAGAAAA TGCATAATGG CCTTTGTTTT GCAGTGGGTT GAAAGGCTTT GAGAATTTGG GTTTGGCAGA TAAATCTGAT GAGTTTTGCT TTTCTGTTTG CTTCCAAGAA CTTAAGGCAG ACAACTTGTT GAACAGAAGT TGTCGCAGCT TACTGTCCAA GAGTATTCCA AAGCATAAGA TAAAAAATCC CTGGAATGCA TTGAGTAAAG CAAAAATAAC ATGCCAAGCC AGATTCTGGC TGTCCACTAT TGTTCCTATT CCAAAGCCCC AGGTGAGCCC TAGCAGAGGG GTCAGAATGA GGAGGCTCTT CCCCACGCGG ATGATGGTGG CCTTGTCATC CCCACTCAGT CTTTCCCCAA CAGTCGGCCT

SEQ ID NO:14

189874

Cluster name: Neuromedin U receptor 2

SequenceID: NM_020167

Sequence: ATGGAAAAAC TTCAGAATGC TTCCTGGATC TACCAGCAGA AACTAGAAGA TCCATTCCAG AAACACCTGA ACAGCACCGA GGAGTATCTG GCCTTCCTCT GCGGACCTCG GCGCAGCCAC TTCTTCCTCC CCGTGTCTGT GGTGTATGTG CCAATTTTTG TGGTGGGGGT CATTGGCAAT GTCCTGGTGT GCCTGGTGAT TCTGCAGCAC CAGGCTATGA AGACGCCCAC CAACTACTAC CTCTTCAGCC TGGCGGTCTC TGACCTCCTG GTCCTGCTCC TTGGAATGCC CCTGGAGGTC TATGAGATGT GGCGCAACTA CCCTTTCTTG TTCGGGCCCG TGGGCTGCTA CTTCAAGACG GCCCTCTTTG AGACCGTGTG CTTCGCCTCC ATCCTCAGCA TCACCACCGT CAGCGTGGAG CGCTACGTGG CCATCCTACA CCCGTTCCGC GCCAAACTGC AGAGCACCCG GCGCCGGGCC CTCAGGATCC TCGGCATCGT CTGGGGCTTC TCCGTGCTCT TCTCCCTGCC CAACACCAGC ATCCATGGCA TCAAGTTCCA CTACTTCCCC AATGGGTCCC TGGTCCCAGG TTCGGCCACC TGTACGGTCA TCAAGCCCAT GTGGATCTAC AATTTCATCA TCCAGGTCAC CTCCTTCCTA TTCTACCTCC TCCCCATGAC TGTCATCAGT GTCCTCTACT ACCTCATGGC ACTCAGACTA AAGAAAGACA AATCTCTTGA GGCAGATGAA GGGAATGCAA ATATTCAAAG ACCCTGCAGA AAATCAGTCA ACAAGATGCT GTTTGTCTTG GTCTTAGTGT TTGCTATCTG TTGGGCCCCG TTCCACATTG ACCGACTCTT CTTCAGCTTT GTGGAGGAGT GGAGTGAATC CCTGGCTGCT GTGTTCAACC TCGTCCATGT GGTGTCAGGT GTCTTCTTCT ACCTGAGCTC AGCTGTCAAC CCCATTATCT ATAACCTACT GTCTCGCCGC TTCCAGGCAG CATTCCAGAA TGTGATCTCT TCTTTCCACA AACAGTGGCA

CTCCCAGCAT GACCCACAGT TGCCACCTGC CCAGCGGAAC ATCTTCCTGA CAGAATGCCA CTTTGTGGAG CTGACCGAAG ATATAGGTCC CCAATTCCCA TGTCAGTCAT CCATGCACAA CTCTCACCTC CCAACAGCCC TCTCTAGTGA ACAGATGTCA AGAACAAACT ATCAAAGCTT CCACTTTAAC AAAACCTGA

SEQ ID NO:15

189876

Cluster name: G protein-coupled receptor Ls189876

SequenceID: ENSMDNA207850

Sequence: ATGAACCAGA CTTTGAATAG CAGTGGGACC GTGGAGTCAG CCCTAAACTA TTCCAGAGGG AGCACAGTGC ACACGGCCTA CCTGGTGCTG AGCTCCCTGG CCATGTTCAC CTGCCTGTGC GGGATGGCAG GCAACAGCAT GGTGATCTGG CTGCTGGGCT TTCGAATGCA CAGGAACCCC TTCTGCATCT ATATCCTCAA CCTGGCGGCA GCCGACCTCC TCTTCCTCTT CAGCATGGCT TCCACGCTCA GCCTGGAAAC CCAGCCCCTG GTCAATACCA CTGACAAGGT CCACGAGCTG ATGAAGAGAC TGATGTACTT TGCCTACACA GTGGGCCTGA GCCTGCTGAC GGCCATCAGC ACCCAGCGCT GTCTCTCTGT CCTCTTCCCT ATCTGGTTCA AGTGTCACCG GCCCAGGCAC CTGTCAGCCT GGGTGTGTGG CCTGCTGTGG ACACTCTGTC TCCTGATGAA CGGGTTGACC TCTTCCTTCT GCAGCAAGTT CTTGAAATTC AATGAAGATC GGTGCTTCAG GGTGGACATG GTCCAGGCCG CCCTCATCAT GGGGGTCTTA ACCCCAGTGA TGACTCTGTC CAGCCTGACC CTCTTTGTCT GGGTGCGGAG GAGCTCCCAG CAGTGGCGGC GGCAGCCCAC ACGGCTGTTC GTGGTGGTCC TGGCCTCTGT CCTGGTGTTC CTCATCTGTT CCCTGCCTCT GAGCATCTAC TGGTTTGTGC TCTACTGGTT GAGCCTGCCG CCCGAGATGC AGGTCCTGTG CTTCAGCTTG TCACGCCTCT CCTCGTCCGT AAGCAGCAGC GCCAACCCCG TCATCTACTT CCTGGTGGGC AGCCGGAGGA GCCACAGGCT GCCCACCAGG TCCCTGGGGA CTGTGCTCCA ACAGGCGCTT CGCGAGGAGC CCGAGCTGGA AGGTGGGGAG ACGCCCACCG TGGGCACCAA TGAGATGGGG GCTTGA

SEQ ID NO:16

189881

Cluster name: G protein-coupled receptor Ls189881

SequenceID: ENSMDNA136950

Sequence: ATGACCCAAC TTGGAAATGA CATTCCCAAG ACCACAAATG ACATTTCCAA GTACCAGGAT GTCTCTATGC CCAGTGCTGG GGCCACACCA GATGCCGAGG CCTCTCCACC CCAGGAGGGC TGCCTCCTCC TCCTAGGTGA CAATGAAGAA TGTACTGCTC AGTCACTGGG CTCAGTGGTC GTCTCTGGGC ATGAGCTGGG TTTCAATGAG CTCAGGAATG GGAAGCATGA CTCTGCCCCT GAGGCCACAT GCCACCTCCA TAGCGGATCT TTTCTTCTGG CTGGAGGGGA AGTCACTTCT TCCCATGAAA CTATTTTATC TATAAATCTC CTCTCCTTGT TGGAGACCAA AGCCCAGCTG CTCCTGCTTG GTGCCCTGGT GGCCTGGGGA CTCAAGGAGT CTCAGAACCT CAAGGTCTGG AGCAGCCCCT ATGTGACCTA CATCCTTAAC CTGGCCACTG TTGATATGGT CAACCTCTCC TGTGTAACTG TGATCCTGCT GGAGAAAATC CTCATGCTGT ATCACCAGGC GGCATTGCAG GTGGCTGTGT TTCTGGATCC TGTCTCCTAT TTCTCCGACA CAGTGGGTCT CTGTCTCCTG GTGGCCATGA GTATTGAGAG CTTTCTCTGT GCCCTCTGTC

CCACCTGGTG CTGCCACCGC CCAGAGCACA CCTCTGCCAT GGCCCTATCT CAAAATATTG TCACATTCAG GGTTAGGACT TTAGCCCGTG AAGTTTGGAT GCCTGGAAGT AAGAGGCAGG TTGATCTCAC AGAGTTGGGC TGCTGCTATG TTCAGGCAGG GGATACAATT TGGGCATTTT ATGTGCCTTT ACCCTGGGCC AACAGTTCCC TTGGAGTGAT TTCATGTCTG CTGGTTTTCA CCATGATTGT GGACCGTTGG TTTTTAAGAG CTGAGGAGGA AGGAACAGGA GTGGAACCAG TTAAAACATC ACAGAGCTCA CTGTTCTTAT CAAGATTCAG CTATTATTCT TGA

189884

SEQ ID NO:17

189883

Cluster name: G protein-coupled receptor Ls 189883

SequenceID: ENSMDNA163742

Sequence: ATGTTGCTGT GCTCTCTGCT TCCCGCCCTT GTGGGATCTC TCTCTGGGGC TGCTGTTTCT GGCCCAATAG GCTGGCGGTT GCCAGGGAAG AGCCCCCGCT TTGACTGTCC AGGGGATGTG GTGGTCAGGG CCAGCTTCTC CATCTTCCAC CTGTACAACA TCACCCTGTT TGATTTCACT GCTCCACCAG CTGGCTTGGA GTCTTCAAGC GTTTCCACCT GGGGCTACTG GGAAGCCCAA GGATTCACAT TTGCCATGGA GGAGATCAAC AGGGACGCCC ACCTGCTCCC CAGCCTCAGG CTGGGCTTCT CCATCCGGAA CTCTGGGCTG GGTATAGTGG CCCTGTGGGA GGCCAAGGTC AGCCCCTCCT CCACACTGGC CAGCCTCAGC GACAGGACCC AGTTCCCATC CTTCTTTCAG ACCCTGCTCA GTCACCTCAC GACCACCCAT GCAGTGGTGC AGCTGATGCT TCACTTCCGA TGGTCTTGGG TGAGCGTCCT GGCGCAGGGG GACGACTTTG AGCTGCAGGG CAGGTCTCTG GTCGTCCAGG AGCTGGGCCA GGCTGGGGTC TGCATTGAAT TCCAACTCTG CATCCCCACC CGGGAGTCCC TGAAGATGAA AAACATCATC TGGCTGATGG AGAACTGTAC GGCCACCATC ATCCTGAAGG AAAGCAAAGT ACACATCGCC TACACAGTGG TCTATGCCAT CGCCCAGGCC CTGGCAGGCT GCAAGCATGG GGACCAGGGG TGTGCCGATG CCTGGGACTT CCAGCCCTGG CTGCTGCTTC GTCCTCTCAA GAACGTGCAT TTCAAGACCC CTGATGGGAC AGAGATCATG TTTGATGCCA ACGGAGATTT AATTACAGAA TTTGATGTTG TCTATGGACA GAAGACCACT GAGGGCTGA

SEQ ID NO:18

LS_ID 189884

Cluster name: G protein-coupled receptor Ls189884

SequenceID: ENSMPRT108574

Sequence: MLAAAFADSN SSSMNVSFAH LHFAGGYLPS DSQDWRTIIP ALLVAVCLVG FVGNLCVIGI LLHNAWKGKP SMIHSLILNL SLADLSLLLF SAPIRATAYS KSVWDLGWFV CKSSDWFIHT CMAAKSLTIV VVAKVCFMYA SDPAKQVSIH NYTIWSVLVA IWTVASLLPL PEWFFSTIRH HEGVEMCLVD VPAVAEEFMS MFGKLYPLLA FGLPLFFASF YFWRAYDQCK KRGTKTQNLR NQIRSKQVTV MLLSIAIISA LLWLPEWVAW LWVWHLKAAG PAPPQGFIAL SQVLMFSISS ANPLIFLVMS EEFREGLKGV WKWMITKKPP TVSESQETPA GNSEGLPDKV PSPESPASIP EKEKPSSPSS GKGKTEKAEI PILPDVEQFW HERDTVPSVQ DNDPIPWEHE DQETGEGV

SEQ ID NO:19

189885

Cluster name: G protein-coupled receptor Ls189885

SequenceID: ENSMDNA178311

Sequence: GGGGCTTCCG AGGTGATCGG GCAGTGTCAG TCTTCAGCCA CTAAGCCGAG AAGATCTGGG AAGGAATCAG TCAGAGAGCC TTGGGCCAGA GTTCCAGGGG CTCTGGGAGT GGGTGTCAGA GAGATTGACC AAACTTTAGG AATTGACACC ATTCTCTGTC ACCATCATGA AAGACTTCTT CAGTCTCATT ACGGAATTCA CAAGTCTTCT TTAATGTCAG TAGGAAATTC ACAAGTCGCA GCTTTGTACC AGCTGAATGT TTATGTTGTT GCTGACACAG TTGGATTAAT TATCAAATCC AATTCAATCC TGGACTCAGT CCAGCCTAAC TATTGCTCAA ATAAACACAT AGAGCTCAGA ACACAAGTTG GTGGAGCTCG GAATCTGAGA GCAAACTCAC CCATGACCTC CAGCTACAAT CAAGAGAGCA GTAGCATGGA GAATGTGTCT GCATTGTCAC TGTTGACTGT GGAGAGTCCC ACGTCCATGT TTGACTATTG TGATGACTCT TTGGAGAGGG TCAAGTCTGC TCTTGACATC TTTTCCATGA TCATCTACAC AGTGACTTTC TTCCTAGGCT TGGCTGGCAA TGGCCTTGTC ATTTGGGTAG TTGGATTCCA CATGTCCTGC ACAGTCAACA CGTGTCTTCC TTCTGACCCT CATCTCCATG GACCACTGAC TTGTGATCCT GTGGCCAATC TAGTCCTGGA ACAATTGCAC ACCAGCAAAG GCAACTCTGG GGCCCTTGAG GACCTGGCTT TTGGCAATTT GTTTCTCTGT TCCCTACTTG ATCTTCAAGG AAACTCGTGG TGGAAAGTGT CACCCTCTTT GTACAACCAG TATGATCTGC AGAATGAAAC TCAAGGAAGT CACCAACTTT GGAAAGAGAT TATCATTCCA TGGCACCAAA CGCTGGTCAC AACAGCCCAC TTTTTCTTTG GCTTCTTTCT CCCTCTGGCT ATCATCACTG GCTACTACAT CCTTGTAGCC TTGAAGTTAA GAGAAAGGCA GCTGGTTAAG TTTAGCTGA

SEQ ID NO:20

189886

Cluster name: G protein-coupled receptor Ls189886

SequenceID: AI659965

Sequence: ACGTATTTTT TATTTTATCA CAACGTCACA GGATGAGACA TTCCCCACTC AAGAAAGTGT ATGTGAAGTT CTGCCTTGAA GAGAGTCAAA TGTCCAAAAC GTAGCCGGAA ATTGGAAGAT GCAAGAAGCA TCAGGAGAGA AGAGGGTCTC TGGGGGACAG CGACTGGGGA GGGCTTGAGG CAGGACTCCA CGCTTATTCC TGTCTGAACC GCCGGAGTGT GGGGGGACGG TGGGGGCAGA GGGAAAGGCC AGGGACTGTC GTCAGGAACA TGCGCTTGGC AGGAAAGCAC GCATTCTATT AGGTTGGTGC ACAAATCACG GCAGAACAGC AGTTTTGCAC CAACCTAATG CTTTACAAAA CACAAAATCA CCCACGTCAA AATGCTCCAT AAATGGCATC AGACTTGGCC GGGCGCAGTG GCTCACGGCT GGGTAATGGT CCACGCTCAC ACAGGCCATG AGGTAGACCC CCCCGTAGGT GTCGGTGTAG AGCACAAACG CCGTCAGCCT GCAGAGCCCC TTGCCGAAAG CCAGCTGGAG CCCAGCACAT AACACACCAC CCTTTCCGGT AAGGCCAGGT GGAACAGCAG TCAG

SEQ ID NO:21

LS_ID 189889

Cluster name: G protein-coupled receptor Ls 189889

SequenceID: ENSMDNA37702

Sequence: ATGCATGTGG GCAGGTATGA AGGACACCCA GACACAGGAG CAGACAACAT GCTGAGAGTG ATATGCTTTG CTTCATTGAA GGTGTCAGGC AGCCGGCAGC ACAGTGGATG TGCAGACCAT GAAGGTGACC CCAAAATCTG CCTGGTGCAC AGCACAAGTG ATGGGGTCTG GGTGGCCAAT GAACATGAAG GGGCAGAGGA AGCTGAGGGC CAAGGAGGAC AGCAGGAGAT AGCTGAGCTG GCAGTTGTTG GCTCGGATGA TGGGAGTGTG GTGGTGTCAG ACGAAGATGC CTAA

SEQ ID NO:22

189895

Cluster name: G protein-coupled receptor GPR61

SequenceID: AF317652

Sequence: ATGGAGTCCT CACCCATCCC CCAGTCATCA GGGAACTCTT CCACTTTGGG GAGGGTCCCT CAAACCCCAG GTCCCTCTAC TGCCAGTGGG GTCCCGGAGG TGGGGCTACG GGATGTTGCT TCGGAATCTG TGGCCCTCTT CTTCATGCTC CTGCTGGACT TGACTGCTGT GGCTGGCAAT GCCGCTGTGA TGGCCGTGAT CGCCAAGACG CCTGCCCTCC GAAAATTTGT CTTCGTCTTC CACCTCTGCC TGGTGGACCT GCTGGCTGCC CTGACCCTCA TGCCCCTGGC CATGCTCTCC AGCCCTGCCC TCTTTGACCA CGCCCTCTTT GGGGAGGTGG CCTGCCGCCT CTACTTGTTT CTGAGCGTGT GCTTTGTCAG CCTGGCCATC CTCTCGGTGT CAGCCATCAA TGTGGAGCGC TACTATTACG TAGTCCACCC CATGCGCTAC GAGGTGCGCA TGACGCTGGG GCTGGTGGCC TCTGTGCTGG TGGGTGTGTG GGTGAAGGCC TTGGCCATGG CTTCTGTGCC AGTGTTGGGA AGGGTCTCCT GGGAGGAAGG AGCTCCCAGT GTCCCCCCAC ACTGTTCACT CCAGTGGAGC CACAGTGCCT ACTGCCAGCT TTTTGTGGTG GTCTTTGCTG TCCTTTACTT TCTGTTGCCC CTGCTCCTCA TACTTCTGGT CTACTGCAGC ATGTTCCGAG TGGCCCGCGT GGCTGCCATG CCAGACGGGC CGCTGCCCAC GTGGATGGAG ACACCCCGGC AACGCTCCGA ATCTCTCAGC AGCCGCTCCA CGATGGTCAC CAGCTCGGGG GCCCCCCAGA CCACCCCACA CCGGACGTTT GGGGGAGGGA AAGCAGCAGT GGTTCTCCTG GCTGTGGGGG GACAGTTCCT GCTCTGTTGG TTGCCCTACT TCTCTTTCCA CCTCTATGTT GCCCTGAGTG CTCAGCCCAT TTCAACTGGG CAGGTGGAGA GTGTGGTCAC CTGGATTGGC TACTTTTGCT TCACTTCCAA CCCTTTCTTC TATGGATGTC TCAACCGGCA GATCCGGGGG GAGCTCAGCA AGCAGTTTGT CTGCTTCTTC AAGCCAGCTC CAGAGGAGGA GCTGAGGCTG CCTAGCCGGG AGGGCTCCAT TGAGGAGAAC TTCCTGCAGT TCCTTCAGGG GACTGGCTGT CCTTCTGAGT CCTGGGTTTC CCGACCCCTA CCCAGCCCCA AGCAGGAGCC ACCTGCTGTT GACTTTCGAA TCCAGGCCAG ATAG

SEQ ID NO:23

189897

Cluster name: G protein-coupled receptor GPR73

SequenceID: AR070166

Sequence: AGCCGCAGAG CGCACAGAAA GGAGGCGCCG AGACAGACAT CACCATGGCA GCCCAGAATG GAAACACCAG TTTCACACCC

AACTTTAATC CACCCCAAGA CCATGCCTCC TCCCTCTCCT
TTAACTTCAG TTATGGTGAT TATGACCTCC CTATGGATGA
GGATGAGGAC ATGACCAAGA CCCGGACCTT CTTCGCAGCC
AAGATCGTCA TTGGCATTGC ACTGGCAGGC ATCATGCTGG
TCTGCGGCAT CGGTAACTTT GTCTTTATCG CTGCCCTCAC
CCGCTATAAG AAGTTGCGCA ACCTCACCAA TCTGCTCATT
GCCAACCTGG CCATCTCCGA CTTCCTGGTG GCCATCATCT
GCTGCCCCTT CGAGATGGAC TACTACGTGG TACGGCAGCT
CTCCTGGGAG CATGGCCACG TGCTCTGTGC CTCCGTCAAC
TACCTGCGCA CCGTCTCCCT CTACGTCTCC ACCAATGCCT
TGCTGGCCAT TGCCATTGAC AGATATCTCG CCATCGTTCA
CCCCTTGAAA CCACGGATGA ATTATCAAAC GGCCTCCTTC
CTGATCGCCT TGGTCTGGAT GGTGTCCATT CTCATTGCCA
TCCCATCGGC TTACTTTGCA ACAGAAACCG TCCTCTTTAT
TGTCAAGAGC CAGGAGAAGA TCTTCTGTGG CCAGATCTGG
CCTGTGGATC AGCAGCTCTA CTACAAGTCC TACTTCCTCT
TCATCTTTGG TGTCGAGTTC GTGGGCCCTG TGGTCACCAT
GACCCTGTGC TATGCCAGGA TCTCCCGGGA GCTCTGGTTC
AAGGCAGTCC CTGGGTTCCA GACGGAGCAG ATTCGCAAGC
GGCTGCGCTG CCGCAGGAAG ACGGTCCTGG TGCTCATGTG
CATTCTCACG GCCTATGTGC TGTGCTGGGC ACCCTTCTAC
GGTTTCACCA TCGTTCGTGA CTTCTTCCCC ACTGTGTTCG
TGAAGGAAAA GCACTACCTC ACTGCCTTCT ACGTGGTCGA
GTGCATCGCC ATGAGCAACA GCATGATCAA CACCGTGTGC
TTCGTGACGG TCAAGAACAA CACCATGAAG TACTTCAAGA
AGATGATGCT GCTGCACTGG CGTCCCTCCC AGCGGGGGAG
CAAGTCCAGT GCTGACCTTG ACCTCAGAAC CAACGGGGTG
CCCACCACAG AAGAAGTGGA CTGTATCAGG CTGAAGTGAC
CCACTGGTGT CACACAATTG AAAACCCCAG TCCAGTACTC
AGAGCATCAC CCACCATCAA CCAAGTTCAT AGGCTGCATG
GGAAATGACA TCTGTGTTCA TGCCTCCCCC GTGCCCTCAA
GAAGCCGAAT GCTGCAAAGT CGTAACATAC AATGAGACTA
GACATGAACC AAATCAGCTG ACATTTACTG ATATCCGCTC
GACACCTACT GTGTCCACAA TCCCCACAAG GAGATTAGAC
ACAAGGAGCA GCAACTGACA TGGACTGAAC ATGTACTGTG
TGCAAACCAC ACCAATGAGA TTAGACGGGG ACAGCAGGAG
CTGACATTTA CTCTTCACCT ACTGTAATCA AAAACACTTG
ATTTGATTAC AATCAAAAAC ATATAAAAAA CATAACAAAG
TAGCAGAAGC TATTGGAGTT TCCAAGCTAT CTCCAGATAT
ATAGATAGTT CACCCTCCAT CTTCCCTAAT TCTGTATCTT
ACCAGTGCAG GAATATCAAA AGGCTATAGG CCAGGCATGA
TGGCTCATGC CTGTAATCCC AGCACTTGGG GAGGCTGAGG
CACGTGGATC ACTTGAGGTC AGGAGTTCAA CCCAGGCTGG
CCAACATGGT GAAACCCTGT CTCTACTAAA AATACAAAAT
TAGCTAGGCG TGGTGGCGGG CGCCTGTAAT CCCAGTTACT
CAGGAGGCTG AAGCAGGAGA ATAGCTTGAA CCTGGGAGTT
GGAGTTTGCA GTGAGCTGAG ATTGCTCCAC TGCACTCCAG
CCTGAGTGAC AGAGTGAGAC TCTGTCTCAG GAAAAAAACA
AACAAACAAA CAACAAAACA ACAACAACAA CAACAACAAC
CAACGGCTAT AGAAGAAGAC TCTTCGACAC AATGGAAATG
TAACGATAAG TTTGTCAGTG CGTGGTTTAC AGCATCATGG
GAGGTGCGTT ACAGCCATCA TACTGAACTT TCCCACCCAC
CTCCTACTGC CTCCCAGGGC ATTCTCTAGG ATTTTGGCTT
CAAGAAAAAA AAAATTCTTA TAGTCAGCCC AGCCTTATGT
GGTTATCCAC AATGGTGTAA TTTCAAAGGA AAGAACCTAA
AAATCACTTT CCCACTGATG CTTGAAAGCT TATCATTTTA
TTTGGGTGGA GATGGGTAAT CCTGAGGTGT CAATTTTTGC
CTCCTCAGTG CAAAGGATTT CAGTGGCTCT GGGGTCAGGG
GGAAAGAGGA CAGAGAAAAA AGTGGAGGTT GCCACTGGCA
ATGAACATAA TCTCTGTGGG CATTTTGCTA AGGACTGGAC

CACTTTCTAG AACACTCCCT CTTTTACAAA AGGAACTCTA CCTAGAATCC AAAGACCTGG GTTCAGGTCC TAACTCTAAG ACTCAAGTCC TAAATTCATG ATGTTTTCTC TCTGTGTCTC AGTTTTGCTT TAATGAAATG GCGATGATGA AAATATCTGC TCTTCATACC TTGCAAGACT GTTGGGAGAG CCCATTGAGG CCATGGTTTG TGAATGTGCT TTTCAACTGT GCACACGATA AGAATGGAGA AGTGATATTG AACAGTTTAT TTGGAGGGAG TTTATTTGGA AACCCCATCC ACTGTGATTT ATTAGAGAAA TACCCACACT TTTTCATCCC TGTTCTTTGG ATGAAAGACT CCTGAAGACT TCACAGTGTA CCTTGTCTAC AGTGGGCCAA AAAGGGATCC CTGTTCTTGG TTATAATCTG GGAAATTTAA CCTCAGATTC TCAGTGACCC CAAGACTCTC AGCATCCCTG CGGTCTTAGA AGTGTTGACA GTCTTCCCTG CATGTTGCAA AATAGCACCC TAGTGCTGCA TAAATATCAC TTCTGAATCT GTTTGTATTA TTATACATTT GTGGTAACTG TAGGTACACG TCTTCATTTC TTCTTGATTC ATTTTGATGT GGTAGCTATG CAAATGGTAC CTGGTTTGGG ACTGACCCAT CCATATTTGA CCAATTCCTA ATTTTTTATA GACAAGGAAT TAATTGTTTG CTTGTTTGAT TGTTTCTATT ATTTGTTGAT TTGTTTCTCT GACTGAAGTT TCAACCAATG TTTCTTTCTA TCACCACCCA GCAGACTCAC CTTCAGCCCA ATCATTGTAC TCTCAGAAAA TGCAGGCCGG CATGGTGGCT CACATCTGTA ATCCCAGCAC TTCGGGAGGC CAAGATGGGC AGATCACCTG AGGTCAGGAG TTCAAGACCA GCCTGGCCAA CATGGCAAAA CCCCATCTCT AGAAAAATAC AGAAATTAGC TGGCGTGGTG GCACATGCCT GTGGTCCCAG CTCCTCAGGA GGCTGAGGCA TGAGAATTGC TTGAACCCCA GAGGCAGAGG TTGCAGTGAA TTGAGATCGC ACCACTGCAC TCCAGCCTGG GTGATAGAGC AAGATTCCAT CTCAAAAGGA AAATAAAAGA AAATGCAAAC ACACTATAAT ATTAGCCTAA GCAAAACTGT TAATTCTGAT TTACAAAAAT TCTTACTTGC TTGGCTTTGA AATGCATTGT GTAATAATGC ATTTCAAAGC CAAGCAAGTA ACAATTTTAG GTTATGTACA

SEQ ID NO:24

189900

Cluster name: Sphingosine 1-phosphate receptor Edg-8

SequenceID: AF317676

Sequence: ATGGAGTCGG GGCTGCTGCG GCCGGCGCCG GTGAGCGAGG TCATCGTCCT GCATTACAAC TACACCGGCA AGCTCCGCGG TGCGCGCTAC CAGCCGGGTG CCGGCCTGCG CGCCGACGCC GTGGTGTGCC TGGCGGTGTG CGCCTTCATC GTGCTAGAGA ATCTAGCCGT GTTGTTGGTG CTCGGACGCC ACCCGCGCTT CCACGCTCCC ATGTTCCTGC TCCTGGGCAG CCTCACGTTG TCGGATCTGC TGGCAGGCGC CGCCTACGCC GCCAACATCC TACTGTCGGG GCCGCTCACG CTGAAACTGT CCCCCGCGCT CTGGTTCGCA CGGGAGGGAG GCGTCTTCGT GGCACTCACT GCGTCCGTGC TGAGCCTCCT GGCCATCGCG CTGGAGCGCA GCCTCACCAT GGCGCGCAGG GGGCCCGCGC CCGTCTCCAG TCGGGGGCGC ACGCTGGCGA TGGCAGCCGC GGCCTGGGGC GTGTCGCTGC TCCTCGGGCT CCTGCCAGCG CTGGGCTGGA ATTGCCTGGG TCGCCTGGAC GCTTGCTCCA CTGTCTTGCC GCTCTACGCC AAGGCCTACG TGCTCTTCTG CGTGCTCGCC TTCGTGGGCA TCCTGGCCGC GATCTGTGCA CTCTACGCGC GCATCTACTG CCAGGTACGC GCCAACGCGC GGCGCCTGCC GGCACGGCCC GGGACTGCGG GGACCACCTC GACCCGGGCG CGTCGCAAGC CGCGCTCGCT GGCCTTGCTG CGCACGCTCA GCGTGGTGCT CCTGGCCTTT GTGGCATGTT GGGGCCCCCT CTTCCTGCTG CTGTTGCTCG ACGTGGCGTG CCCGGCGCGC

ACCTGTCCTG TACTCCTGCA GGCCGATCCC TTCCTGGGAC TGGCCATGGC CAACTCACTT CTGAACCCCA TCATCTACAC GCTCACCAAC CGCGACCTGC GCCACGCGCT CCTGCGCCTG GTCTGCTGCG GACGCCACTC CTGCGGCAGA GACCCGAGTG GCTCCCAGCA GTCGGCGAGC GCGGCTGAGG CTTCCGGGGG CCTGCGCCGC TGCCTGCCCC CGGGCCTTGA TGGGAGCTTC AGCGGCTCGG AGCGCTCATC GCCCCAGCGC GACGGGCTGG ACACCAGCGG CTCCACAGGC AGCCCCGGTG CACCCACAGC CGCCCGGACT CTGGTATCAG AACCGGCTGC AGACTGA

SEQ ID NO:25

189901

Cluster name: G protein-coupled receptor Ls189901

SequenceID: E31720

Sequence: GACTATCCTC CCACTTCAGG GTTTCTCTGG GCTTCCATCT TGCCCCTGCT GAGCCCTGCT TCCTCCTCTA CCAGCAGCAC AACCCCCAGG CTGGGCTCAG AGACCTCATG TGGTGGGATC ACTCAGTACC CCGAGGCGGA GGGAAGGAGG GAGGGCTGCA GGGTTCCCCT TGGCCTGCAA ACAGGAACAC AGGGTGTTTC TCAGTGGCTG CGAGAATGCT GATGAAAACC CCAGGATGTT GTGTCACCGT GGTGGCCAGC TGATAGTGCC AATCATCCCA CTTTGCCCTG AGCACTCCTG CAGGGGTAGA AGACTCCAGA ACCTTCTCTC AGGCCCATGG CCCAAGCAGC CCATGGAACT TCATAACCTG AGCTCTCCAT CTCCCTCTCT CTCCTCCTCT GTTCTCCCTC CCTCCTTCTC TCCCTCACCC TCCTCTGCTC CCTCTGCCTT TACCACTGTG GGGGGGTCCT CTGGAGGGCC CTGCCACCCC ACCTCTTCCT CGCTGGTGTC TGCCTTCCTG GCACCAATCC TGGCCCTGGA GTTTGTCCTG GGCCTGGTGG GGAACAGTTT GGCCCTCTTC ATCTTCTGCA TCCACACGCG GCCCTGGACC TCCAACACGG TGTTCCTGGT CAGCCTGGTG GCCGCTGACT TCCTCCTGAT CAGCAACCTG CCCCTCCGCG TGGACTACTA CCTCCTCCAT GAGACCTGGC GCTTTGGGGC TGCTGCCTGC AAAGTCAACC TCTTCATGCT GTCCACCAAC CGCACGGCCA GCGTTGTCTT CCTCACAGCC ATCGCACTCA ACCGCTACCT GAAGGTGGTG CAGCCCCACC ACGTGCTGAG CCGTGCTTCC GTGGGGGCAG CTGCCCGGGT GGCCGGGGGA CTCTGGGTGG GCATCCTGCT CCTCAACGGG CACCTGCTCC TGAGCACCTT CTCCGGCCCC TCCTGCCTCA GCTACAGGGT GGGCACGAAG CCCTCGGCCT CGCTCCGCTG GCACCAGGCA CTGTACCTGC TGGAGTTCTT CCTGCCACTG GCGCTCATCC TCTTTGCTAT TGTGAGCATT GGGCTCACCA TCCGGAACCG TGGTCTGGGC GGGCAGGCAG GCCCGCAGAG GGCCATGCGT GTGCTGGCCA TGGTGGTGGC CGTCTACACC ATCTGCTTCT TGCCCAGCAT CATCTTTGGC ATGGCTTCCA TGGTGGCTTT CTGGCTGTCC GCCTGCCGCT CCCTGGACCT CTGCACACAG CTCTTCCATG GCTCCCTGGC CTTCACCTAC CTCAACAGTG TCCTGGACCC CGTGCTCTAC TGCTTCTCTA GCCCCAACTT CCTCCACCAG AGCCGGGCCT TGCTGGGCCT CACGCGGGGC CGGCAGGGCC CAGTGAGCGA CGAGAGCTCC TACCAACCCT CCAGGCAGTG GCGCTACCGG GAGGCCTCTA GGAAGGCGGA GGCCATAGGG AAGCTGAAAG TGCAGGGCGA GGTCTCTCTG GAAAAGGAAG GCTCCTCCCA GGGCTGAGGG CCAGCTGCAG GGCTGCAGCG CTGTGGGGGT AAGGGCTGCC GCGCTCTGGC CTGGAGGGAC AAGGCCAGCA CACGGTGCCT CAAC

SEQ ID NO:26

190188

Cluster name: G protein-coupled receptor LGR6

SequenceID: AB049405

Sequence: GCCACTGCCA GGAGGACGGC ATCATGCTGT CTGCCGACTG CTCTGAGCTC GGGCTGTCCG CCGTTCCGGG GGACCTGGAC CCCCTGACGG CTTACCTGGA CCTCAGCATG AACAACCTCA CAGAGCTTCA GCCTGGCCTC TTCCACCACC TGCGCTTCTT GGAGGAGCTG CGTCTCTCTG GGAACCATCT CTCACACATC CCAGGACAAG CATTCTCTGG TCTCTACAGC CTGAAAATCC TGATGCTGCA GAACAATCAG CTGGGAGGAA TCCCCGCAGA GGCGCTGTGG GAGCTGCCGA GCCTGCAGTC GCTGCGCCTA GATGCCAACC TCATCTCCCT GGTCCCGGAG AGGAGCTTTG AGGGGCTGTC CTCCCTCCGC CACCTCTGGC TGGACGACAA TGCACTCACG GAGATCCCTG TCAGGGCCCT CAACAACCTC CCTGCCCTGC AGGCCATGAC CCTGGCCCTC AACCGCATCA GCCACATCCC CGACTACGCG TTCCAGAATC TCACCAGCCT TGTGGTGCTG CATTTGCATA ACAACCGCAT CCAGCATCTG GGGACCCACA GCTTCGAGGG GCTGCACAAT CTGGAGACAC TAGACCTGAA TTATAACAAG CTGCAGGAGT TCCCTGTGGC CATCCGGACC CTGGGCAGAC TGCAGGAACT GGGGTTCCAT AACAACAACA TCAAGGCCAT CCCAGAAAAG GCCTTCATGG GGAACCCTCT GCTACAGACG ATACACTTTT ATGATAACCC AATCCAGTTT GTGGGAAGAT CGGCATTCCA GTACCTGCCT AAACTCCACA CACTATCTCT GAATGGTGCC ATGGACATCC AGGAGTTTCC AGATCTCAAA GGCACCACCA GCCTGGAGAT CCTGACCCTG ACCCGCGCAG GCATCCGGCT GCTCCCATCG GGGATGTGCC AACAGCTGCC CAGGCTCCGA GTCCTGGAAC TGTCTCACAA TCAAATTGAG GAGCTGCCCA GCCTGCACAG GTGTCAGAAA TTGGAGGAAA TCGGCCTCCA ACACAACCGC ATCTGGGAAA TTGGAGCTGA CACCTTCAGC CAGCTGAGCT CCCTGCAAGC CCTGGATCTT AGCTGGAACG CCATCCGGTC CATCCACCCT GAGGCCTTCT CCACCCTGCA CTCCCTGGTC AAGCTGGACC TGACAGACAA CCAGCTGACC ACACTGCCCC TGGCTGGACT TGGGGGCTTG ATGCATCTGA AGCTCAAAGG GAACCTTGCT CTCTCCCAGG CCTTCTCCAA GGACAGTTTC CCAAAACTGA GGATCCTGGA GGTGCCTTAT GCCTACCAGT GCTGTCCCTA TGGGATGTGT GCCAGCTTCT TCAAGGCCTC TGGGCAGTGG GAGGCTGAAG ACCTTCACCT TGATGATGAG GAGTCTTCAA AAAGGCCCCT GGGCCTCCTT GCCAGACAAG CAGAGAACCA CTATGACCAG GACCTGGATG AGCTCCAGCT GGAGATGGAG GACTCAAAGC CACACCCCAG TGTCCAGTGT AGCCCTACTC CAGGCCCCTT CAAGCCCTGT GAGTACCTCT TTGAAAGCTG GGGCATCCGC CTGGCCGTGT GGGCCATCGT GTTGCTCTCC GTGCTCTGCA ATGGACTGGT GCTGCTGACC GTGTTCGCTG GCGGGCCTGC CCCCCTGCCC CCGGTCAAGT TTGTGGTAGG TGCGATTGCA GGCGCCAACA CCTTGACTGG CATTTCCTGT GGCCTTCTAG CCTCAGTCGA TGCCCTGACC TTTGGTCAGT TCTCTGAGTA CGGAGCCCGC TGGGAGACGG GGCTAGGCTG CCGGGCCACT GGCTTCCTGG CAGTACTTGG GTCGGAGGCA TCGGTGCTGC TGCTCACTCT GGCCGCAGTG CAGTGCAGCG TCTCCGTCTC CTGTGTCCGG GCCTATGGGA AGTCCCCCTC CCTGGGCAGC GTTCGAGCAG GGGTCCTAGG CTGCCTGGCA CTGGCAGGGC TGGCCGCCGC ACTGCCCCTG GCCTCAGTGG GAGAATACGG GGCCTCCCCA CTCTGCCTGC CCTACGCGCC ACCTGAGGGT CAGCCAGCAG CCCTGGGCTT CACCGTGGCC CTGGTGATGA TGAACTCCTT CTGTTTCCTG GTCGTGGCCG GTGCCTACAT CAAACTGTAC TGTGACCTGC CGCGGGGCGA CTTTGAGGCC GTGTGGGACT GCGCCATGGT GAGGCACGTG GCCTGGCTCA TCTTCGCAGA CGGGCTCCTC TACTGTCCCG TGGCCTTCCT CAGCTTTGCC TCCATGCTGG

GCCTCTTCCC TGTCACGCCC GAGGCCGTCA AGTCTGTCCT GCTGGTGGTG CTGCCCCTGC CTGCCTGCCT CAACCCACTG CTGTACCTGC TCTTCAACCC CCACTTCCGG GATGACCTTC GGCGGCTTCG GCCCCGCGCA GGGGACTCAG GGCCCCTAGC CTATGCTGCG GCCGGGGAGC TGGAGAAGAG CTCCTGTGAT TCTACCCAGG CCCTGGTAGC CTTCTCTGAT GTGGATCTCA TTCTGGAAGC TTCTGAAGCT GGGCGGCCCC CTGGGCTGGA GACCTATGGC TTCCCCTCAG TGACCCTCAT CTCCTGTCAG CAGCCAGGGG CCCCCAGGCT GGAGGGCAGC CATTGTGTAG AGCCAGAGGG GAACCACTTT GGGAACCCCC AACCCTCCAT GGATGGAGAA CTGCTGCTGA GGGCAGAGGG ATCTACGCCA GCAGGTGGAG GCTTGTCAGG GGGTGGCGGC TTTCAGCCCT CTGGCTTGGC CTTTGCTTCA CACGTGTAAA TATCCCTCCC CATTCTTCTC TTCCCCTCTC TTCCCTTTCC TCTCTCCCCC TCGGTGAATG ATGGCTGCTT CTAAAACAAA TACAACCAAA ACTCAGCAGT GTGATCTATA GCAGGATGGC CCAGTACCTG GCTCCACTGA TCACCTCTCT CCTGTGACCA TCACCAACGG GTGCCTCTTG GCCTGGCTTT CCCTTGGCCT TCCTCAGCTT

SEQ ID NO:27

190411

Cluster name: G protein-coupled receptor Ls190411

SequenceID: AF305409

Sequence: CCACAAGGAG TAGTTGGGAG ATACAGGGGC ATGGCCACCA CAAGCAGAAT AATTTTCGGG ATATTTTGTA GAAGATGGGG TTTTGCCACA TTGCCCAGGC TGGTCTCGAA CTGGGTGGGA TCAAACGATC CAACCGCGTT GGCCTCCAGA GTGTTGGGAT TACAGGTGTG AGCCACCAAG CATGGAATAG GCTTCTTTAA ACATTGAATA GTATTCCTTT GGTAGATGAA GGAGGATGAG ATAGCACGAG AGGGCAAAGA TGCAGCCAAG TAACCCAGTG CTGGAGCCCA CGATGGAGAA GATCTCACGG CCACTCTGGC CTTGCCCTGG GTGCTTTAGT AACTCGGGAG GAAGGCCACC CAGACACTGC AGGACACCAG CATGCTGAAG GTCAGGAACT TGACTTATTG AAGGTGTCAG GCAGGTTCCT TGCCAGAAAG GCTACAGCAA GGGACCCTAA AACCAAGAAG CCCAAGTAGC CCAAGACAGA GTAGAAGGCA GTGACGGAGC CCTCATTACA CTGGATAATG ATGTAGCCAG GCATGAACTG AGGGTCCTTG TTTACGAAGG GAGGCTCTGT CCCCAGCCAG ATTCCACAGA GGGTC

SEQ ID NO:28

190414

Cluster name: G protein-coupled receptor Ls190414

SequenceID: AX080495

Sequence: GCCTGCAACC TGTCYCACGC CCTCTGGCTG TTGCCATGAC GTCCACCTGC ACCAACAGCA CGCGCGAGAG TAACAGCAGC CACACGTGCA TGCCCCTCTC CAAAATGCCC ATCAGCCTGG CCCACGGCAT CATCCGCTCA ACCGTGCTGG TTATCTTCCT CGCCGCCTCT TTCGTCGGCA ACATAGTGCT GGCGCTAGTG TTGCAGCGCA AGCCGCAGCT GCTGCAGGTG ACCAACCGTT TTATCTTTAA CCTCCTCGTC ACCGACCTGC TGCAGATTTC GCTCGTGGCC CCCTGGGTGG TGGCCACCTC TGTGCCTCTC TTCTGGCCCC TCAACAGCCA CTTCTGCACG GCCCTGGTTA GCCTCACCCA CCTGTTCGCC TTCGCCAGCG TCAACACCAT TGTCTTGGTG TCAGTGGATC GCTACTTGTC CATCATCCAC

CCTCTCTCCT ACCCGTCCAA GATGACCCAG CGCCGCGGTT
ACCTGCTCCT CTATGGCACC TGGATTGTGG CCATCCTGCA
GAGCACTCCT CCACTCTACG GCTGGGGCCA GGCTGCCTTT
GATGAGCGCA ATGCTCTCTG CTCCATGATC TGGGGGGCCA
GCCCCAGCTA CACTATTCTC AGCGTGGTGT CCTTCATCGT
CATTCCACTG ATTGTCATGA TTGCCTGCTA CTCCGTGGTG
TTCTGTGCAG CCCGGAGGCA GCATGCTCTG CTGTACAATG
TCAAGAGACA CAGCTTGGAA GTGCGAGTCA AGGACTGTGT
GGAGAATGAG GATGAAGAGG GAGCAGAGAA GAAGGAGGAG
TTCCAGGATG AGAGTGAGTT TCGCCGCCAG CATGAAGGTG
AGGTCAAGGC CAAGGAGGGC AGAATGGAAG CCAAGGACGG
CAGCCTGAAG GCCAAGGAAG GAAGCACGGG GACCAGTGAG
AGTAGTGTAG AGGCCAGGGG CAGCGAGGAG GTCAGAGAGA
GCAGCACGGT GGCCAGCGAC GGCAGCATGG AGGGTAAGGA
AGGCAGCACC AAAGTTGAGG AGAACAGCAT GAAGGCAGAC
AAGGGTCGCA CAGAGGTCAA CCAGTGCAGC ATTGACTTGG
GTGAAGATGG CATGGAGTTT GGTGAAGACG ACATCAATTT
CAGTGAGGAT GACGTCGAGG CAGTGAACAT CCCGGAGAGC
CTCCCACCCA GTCGTCGTAA CAGCAACAGC AACCCTCCTC
TGCCCAGGTG CTACCAGTGC AAAGCTGCTA AAGTGATCTT
CATCATCATT TTCTCCTATG TGCTATCCCT GGGGCCCTAC
TGCTTTTTAG CAGTCCTGGC CGTGTGGGTG GATGTCGAAA
CCCAGGTACC CCAGTGGGTG ATCACCATAA TCATCTGGCT
TTTCTTCCTG CAGTGCTGCA TCCACCCCTA TGTCTATGGC
TACATGCACA AGACCATTAA GAAGGAAATC CAGGACATGC
TGAAGAAGTT CTTCTGCAAG GAAAAGCCCC CGAAAGAAGA
TAGCCACCCA GACCTGCCCG GAACAGAGGG TGGGACTGAA
GGCAAGATTG TCCCTTCCTA CGATTCTGCT ACTTTTCCTT
GAAGTTAGTT CTAAGGCAAA CCTTGAAAAT CAGTCCTTCA
GCCACAGCTA TTTAGAGCTT TAAAACTACC AGGTTCAATC
ACTGGTTATG CTTTCTGTG

SEQ ID NO:29

190418

Cluster name: G protein-coupled receptor EX33 (GPR84)

SequenceID: NM_020370

Sequence: TAACTGTCCA CCAGAAAGGA CTGCTCTTTG GGTGAGTTGA
ACTTCTTCCA TTATAGAAAG AATTGAAGGC TGAGAAACTC
AGCCTCTATC ATGTGGAACA GCTCTGACGC CAACTTCTCC
TGCTACCATG AGTCTGTGCT GGGCTATCGT TATGTTGCAG
TTAGCTGGGG GGTGGTGGTG GCTGTGACAG GCACCGTGGG
CAATGTGCTC ACCCTACTGG CCTTGGCCAT CCAGCCCAAG
CTCCGTACCC GATTCAACCT GCTCATAGCC AACCTCACAC
TGGCTGATCT CCTCTACTGC ACGCTCCTTC AGCCCTTCTC
TGTGGACACC TACCTCCACC TGCACTGGCG CACCGGTGCC
ACCTTCTGCA GGGTATTTGG GCTCCTCCTT TTTGCCTCCA
ATTCTGTCTC CATCCTGACC CTCTGCCTCA TCGCACTGGG
ACGCTACCTC CTCATTGCCC ACCCTAAGCT TTTTCCCCAA
GTTTTCAGTG CCAAGGGGAT AGTGCTGGCA CTGGTGAGCA
CCTGGGTTGT GGGCGTGGCC AGCTTTGCTC CCCTCTGGCC
TATTTATATC CTGGTACCTG TAGTCTGCAC CTGCAGCTTT
GACCGCATCC GAGGCCGGCC TTACACCACC ATCCTCATGG
GCATCTACTT TGTGCTTGGG CTCAGCAGTG TTGGCATCTT
CTATTGCCTC ATCCACCGCC AGGTCAAACG AGCAGCACAG
GCACTGGACC AATACAAGTT GCGACAGGCA AGCATCCACT
CCAACCATGT GGCCAGGACT GATGAGGCCA TGCCTGGTCG
TTTCCAGGAG CTGGACAGCA GGTTAGCATC AGGAGGACCC
AGTGAGGGGA TTTCATCTGA GCCAGTCAGT GCTGCCACCA

CCCAGACCCT GGAAGGGGAC TCATCAGAAG TGGGAGACCA GATCAACAGC AAGAGAGCTA AGCAGATGGC AGAGAAAAGC CCTCCAGAAG CATCTGCCAA AGCCCAGCCA ATTAAAGGAG CCAGAAGAGC TCCGGATTCT TCATCGGAAT TTGGGAAGGT GACTCGAATG TGTTTTGCTG TGTTCCTCTG CTTTGCCCTG AGCTACATCC CCTTCTTGCT GCTCAACATT CTGGATGCCA GAGTCCAGGC TCCCCGGGTG GTCCACATGC TTGCTGCCAA CCTCACCTGG CTCAATGGTT GCATCAACCC TGTGCTCTAT GCAGCCATGA ACCGCCAATT CCGCCAAGCA TATGGCTCCA TTTTAAAAAG AGGGCCCCGG AGTTTCCATA GGCTCCATTA GAACTGTGAC CCTAGTCACC AGAATTCAGG ACTGTCTCCT CCAGGACCAA AGTGGCCAGG TAATAGGAGA ATAGGTGAAA TAACACATGT GGGCATTTTC ACAACAATCT CTCCCCAGCC TCCCAAATCA AGTCTCTCCA TCACTTGATC AATGTTTCAG CCCTAGACTG CCCAAGGAGT ATTATTAATT ATTAATAAAT GAATTCTGTG CTTTTAAAAA AAAAAAAATA AAAAAAGAAA AAAAAAAAAA AAAAAAAAAA AAAAAA

SEQ ID NO:30

190419

Cluster name: G protein-coupled receptor Ls190419

SequenceID: AJ303165

Sequence: CTTTGCTTCA GAGCTAAACC AGTTTTTCTT CTCTCCACAG CAAATATCTT GACAGTGATC ATCCTCTCCC AGCTGGTGGC AAGAAGACAG AAGTCCTCCT ACAACTATCT CTTGGCACTC GCTGCTGCCG ACATCTTGGT CCTCTTTTTC ATAGTGTTTG TGGACTTCCT GTTGGAAGAT TTCATCTTGA ACATGCAGAT GCCTCAGGTC CCCGACAAGA TCATAGAAGT GCTGGAATTC TCATCCATCC ACACCTCCAT ATGGATTACT GTACCGTTAA CCATTGACAG GTATATCGCT GTCTGCCACC CGCTCAAGTA CCACACGGTC TCATACCCAG CCCGCACCCG GAAAGTCATT GTAAGTGTTT ACATCACCTG CTTCCTGACC AGCATCCCCT ATTACTGGTG GCCCAACATC TGGACTGAAG ACTACATCAG CACCTCTGTG CATCACGTCC TCATCTGGAT CCACTGCTTC ACCGTCTACC TGGTGCCCTG CTCCATCTTC TTCATCTTGA ACTCAATCAT TGTGTACAAG CTCAGGAGGA AGAGCAATTT TCGTCTCCGT GGCTACTCCA CGGGGAAGAC CACCGCCATC TTGTTCACCA TTACCTCCAT CTTTGCCACA CTTTGGGCCC CCCGCATCAT CATGATTCTT TACCACCTCT ATGGGGCGCC CATCCAGAAC CGCTGGCTGG TGCACATCAT GTCCGACATT GCCAACATGC TAGCCCTTCT GAACACAGCC ATCAACTTCT TCCTCTACTG CTTCATCAGC AAGCGGTTCC GCACC

SEQ ID NO:31

190427

Cluster name: Cysteinyl leukotriene CysLT2 receptor

SequenceID: NM_020377

Sequence: AAGTTCTCTA AGTTTGAAGC GTCAGCTTCA ACCAAACAAA TTAATGGCTA TTCTACATTC AAAAATCAGG AAATTTAAAT TTATTATGAA ATGTAATGCA GCATGTAGTA AAGACTTAAC CAGTGTTTTA AAACTCAACT TTCAAAGAAA AGATAGTATT GCTCCCTGTT TCATTAAAAC CTAGAGAGAT GTAATCAGTA AGCAAGAAGG AAAAAGGGAA ATTCACAAAG TAACTTTTTG TGTCTGTTTC TTTTTAACCC AGCATGGAGA GAAAATTTAT

GTCCTTGCAA CCATCCATCT CCGTATCAGA AATGGAACCA
AATGGCACCT TCAGCAATAA CAACAGCAGG AACTGCACAA
TTGAAAACTT CAAGAGAGAA TTTTTCCCAA TTGTATATCT
GATAATATTT TTCTGGGGAG TCTTGGGAAA TGGGTTGTCC
ATATATGTTT TCCTGCAGCC TTATAAGAAG TCCACATCTG
TGAACGTTTT CATGCTAAAT CTGGCCATTT CAGATCTCCT
GTTCATAAGC ACGCTTCCCT TCAGGGCTGA CTATTATCTT
AGAGGCTCCA ATTGGATATT TGGAGACCTG GCCTGCAGGA
TTATGTCTTA TTCCTTGTAT GTCAACATGT ACAGCAGTAT
TTATTTCCTG ACCGTGCTGA GTGTTGTGCG TTTCCTGGCA
ATGGTTCACC CCTTTCGGCT TCTGCATGTC ACCAGCATCA
GGAGTGCCTG GATCCTCTGT GGGATCATAT GGATCCTTAT
CATGGCTTCC TCAATAATGC TCCTGGACAG TGGCTCTGAG
CAGAACGGCA GTGTCACATC ATGCTTAGAG CTGAATCTCT
ATAAAATTGC TAAGCTGCAG ACCATGAACT ATATTGCCTT
GGTGGTGGGC TGCCTGCTGC CATTTTTCAC ACTCAGCATC
TGTTATCTGC TGATCATTCG GGTTCTGTTA AAAGTGGAGG
TCCCAGAATC GGGGCTGCGG GTTTCTCACA GGAAGGCACT
GACCACCATC ATCATCACCT TGATCATCTT CTTCTTGTGT
TTCCTGCCCT ATCACACACT GAGGACCGTC CACTTGACGA
CATGGAAAGT GGGTTTATGC AAAGACAGAC TGCATAAAGC
TTTGGTTATC ACACTGGCCT TGGCAGCAGC CAATGCCTGC
TTCAATCCTC TGCTCTATTA CTTTGCTGGG GAGAATTTTA
AGGACAGACT AAAGTCTGCA CTCAGAAAAG GCCATCCACA
GAAGGCAAAG ACAAAGTGTG TTTTCCCTGT TAGTGTGTGG
TTGAGAAAGG AAACAAGAGT ATAAGGAGCT CTTAGATGAG
ACCTGTTCTT GTATCCTTGT GTCCATCTTC ATTCACTCAT
AGTCTCCAAA TGACTTTGTA TTTACATCAC TCCCAACAAA
TGTTGATTCT TAATATTTAG TTGACCATTA CTTTTGTTAA
TAAGACCTAC TTCAAAAATT TTATTCAGTG TATTTTCAGT
TGTTGAGTCT TAATGAGGGA TACAGGAGGA AAAATCCCTA
CTAGAGTCCT GTGGGCTGAA ATATCAGACT GGGAAAAAAT
GCAAAGCACA TTGGATCCTA CTTTTCTTCA GATATTGAAC
CAGATCTCTG GCCCATCAGG CTTTCTAAAT TCTTCAAAAG
AGCCACAACT TCCCCAGCTT CTCCAGCTCC CCTGTCCTCT
TCAATCCCTT GAGATATAGC AACTAACGAC GCTACTGGAA
GCCCCAGAGC AGAAAAGAAG CACATCCTAA GATTCAGGGA
AAGACTAACT GTGAAAAGGA AGGCTGTCCT ATAACAAAGC
AGCATCAAGT CCCAAGTAAG GACAGTGAGA GAAAAGGGGG
AGAAGGATTG GAGCAAAAGA GAACTGGCAA TAAGTAGGGG
AAGGAAGAAT TTCATTTTGC ATTGGGAGAG AGGTTCTAAC
ACACTGAAGG CAACCCTATT TCTACTGTTT CTCTCTTGCC
AGGGTATTAG GAAGGACAGG AAAAGTAGGA GGAGGATCTG
GGGCATTGCC CTAGGAAATG AAAGAATTGT GTATAGAATG
GAAGGGGGAT CATCAAGGAC ATGTATCTCA AATTTTCTTT
GAGATGCAGG TTAGTTGACC TTGCTGCAGT TCTCCTTCCC
ATTAATTCAT TGGGATGGAA GCCAAAAATA AAAGAGGTGC
CTCTGAGGAT TAGGGTTGAG CACTCAAGGG AAAGATGGAG
TAGAGGGCAA ATAGCAAAAG TTGTTGCACT CCTGAAATTC
TATTAACATT TCCGCAGAAG ATGAGTAGGG AGATGCTGCC
TTCCCTTTTG AGATAGTGTA GAAAAACACT AGATAGTGTG
AGAGGTTCCT TTCTGTCCAT TGAAACAAGG CTAAGGATAC
TACCAACTAC TATCACCATG ACCATTGTAC TGACAACAAT
TGAATGCAGT CTCCCTGCAG GGCAGATTAT GCCAGGCACT
TTACATTTGT TGATCCCATT TGACATTCAC ACCAAAGCTC
TGAGTTCCAT TTTACAGCTG AAGAAATTGA AGCTTAGAGA
AATTAAGAAG CTTGTTTAAG TTTACACAGC TAGTAAGAGT
TTTAAAAATC TCTGTGCAGA AGTGTTGGCT GGGTGCTCTC
CCCACCACTA CCCTTGTAAA CTTCCAGGAA GATTGGTTGA
AAGTCTGAAT AAAAGCTGTC CTTTCCTACC AATTTCCTCC

CCCTCCTCAC TCTCACAAGA AAACCAAAAG TTTCTCTTCA

SEQ ID NO:32

190428

Cluster name: G protein-coupled receptor Ls190428

SequenceID: AX100250

Sequence: GAGCAGAAAT TCGGCACGAG GAAAAATCTG AAATCTGAAA TGCTCCAAAA TCCTAAACTT TTTGAGTGCT GACATTATGC CACAAATGGA AAATTTCATA CCTGACCTTA TGTGAGTTGC AGTCAAAACA CAGGTGCACA ACACCCAGTT CATGCAACAT CCCCAATGGG AAAAAAGACC CCCCCAGCTC TCTTCTGCTG CAGTTTTTCT GCTCACACCT GGATTCCCCA TGCATTCCCA CAAAAAGTAA TTAAATGGCA TGCGTGCAGG CTGGACACGC CAACAACAGG TTTCCCACAA TGCCCCACAT GGGCGAAGAC CTGTGTGCAT TACTCATTGC ATTTTTTTGC TTATTCTCTG CTGTGTGGTA TAAATATATT GTTGAAAATG TCAAAAAGAC CTAAAGATAC CCCTGTGAAT ATCAGTGATA AGAAAAAGAG GAAGCATTTA TGTTTATCTA TAGCACAGAA AGTCAAGTTG TTGGAGAAAC TGGACAGTGG TGTAAGTGTG AAACATCTTA CAGAAGAGTA TGGTGTTGGA ATGACCACCA TATATGACCT GAAGAAACAG AAGGATAAAC TGTTGAAGTT TTATGCTGAA AGTGATGAGC AGATATTAAT GAAAAATAGA AAAACACTTC ATAAAGCTAA AAATGAAGAT CTTGATCGTG TATTGAAAGA GTGGATCCGT CAGCGTCGCA GTGAACACAT GCCACTTAAT GGTATGCTGA TCATGAAACA AGCAAAGATA TATCACAATG AACTAAAAAT TGAGGGGAAC TGTGAATATT CAACAGGCTG GTTGCAGAAA TTTAAGAAAA GACATGGCAT TAAATTTTTA AAGACTTGTG GCAATAAAGC ATCTGCTGGT CATGAAGCAA CAGAGAAGTT TACTGGCAAT TTCAGTAATG ATGATGAACA AGATGGTAAC TTTGAAGGAT TCAGTATGTC AAGTGAGAAA AAAATAATGT CTGACCTCCT TACATATACA AAAAATATAC ATCCAGAGAC TGTCAGTAAG CTGGAAGAAG AGGATATCAA AGATGTTTTT AACAGTAATA ATGAGGCTCC AGTTGTTCAT TCATTGTCCA ATGGTGAAGT AACAAAAATG GTTCTGAATC AAGATGATCA TGATGATAAT GATAATGAAG ATGATGTTAA CACTGCAGAA AAAGTGCCTA TAGACGACAT GGTAAAAATG TGTGATGGGC TTATTAAAGG ACTAGAGCAG CATGCATTCA TAACAGAGCA AGAAATCATG TCAGTTTATA AAATCAAAGA GAGACTTCTA AGACAAAAAG CATCATTAAT GAGGCAGATG ACTCTGAAAG AAACATTTAA AAAAGCCATC CAGAGGAATG CTTCTTCCTC TCTACAGGAC CCACTTCTTG GTCCCTCAAC TGCTTCTGAT GCTTCTTCTC ACCTAAAAAT AAAATAAAAT ACAGTGTACA GTAACCTTTT AGTCAAAACA GCATCATACT TGGAAACTGA AAGCC

SEQ ID NO:33

190437

Cluster name: G protein-coupled receptor C5L2

SequenceID: NM_018485

Sequence: CCTGTGTGCC ACGTGCTGGA CAAATCTTAA CTCCTCAAGG ACTCCCAAAA CCAGAGACAC CAGGAGCCTG AATGGGGAAC GATTCTGTCA GCTACGAGTA TGGGGATTAC AGCGACCTCT CGGACCGCCC TGTGGACTGC CTGGATGGCG CCTGCCTGGC CATCGACCCG CTGCGCGTGG CCCCGCTCCC ACTGTATGCC

GCCATCTTCC TGGTGGGGGT GCCGGGCAAT GCCATGGTGG CCTGGGTGGC TGGGAAGGTG GCCCGCCGGA GGGTGGGTGC CACCTGGTTG CTCCACCTGG CCGTGGCGGA TTTGCTGTGC TGTTTGTCTC TGCCCATCCT GGCAGTGCCC ATTGCCCGTG GAGGCCACTG GCCGTATGGT GCAGTGGGCT GTCGGGCGCT GCCCTCCATC ATCCTGCTGA CCATGTATGC CAGCGTCCTG CTCCTGGCAG CTCTCAGTGC CGACCTCTGC TTCCTGGCTC TCGGGCCTGC CTGGTGGTCT ACGGTTCAGC GGGCGTGCGG GGTGCAGGTG GCCTGTGGGG CAGCCTGGAC ACTGGCCTTG CTGCTCACCG TGCCCTCCGC CATCTACCGC CGGCTGCACC AGGAGCACTT CCCAGCCCGG CTGCAGTGTG TGGTGGACTA CGGCGGCTCC TCCAGCACCG AGAATGCGGT GACTGCCATC CGGTTTCTTT TTGGCTTCCT GGGGCCCCTG GTGGCCGTGG CCAGCTGCCA CAGTGCCCTC CTGTGCTGGG CAGCCCGACG CTGCCGGCCG CTGGGCACAG CCATTGTGGT GGGGTTTTTT GTCTGCTGGG CACCCTACCA CCTGCTGGGG CTGGTGCTCA CTGTGGCGGC CCCGAACTCC GCACTCCTGG CCAGGGCCCT GCGGGCTGAA CCCCTCATCG TGGGCCTTGC CCTCGCTCAC AGCTGCCTCA ATCCCATGCT CTTCCTGTAT TTTGGGAGGG CTCAACTCCG CCGGTCACTG CCAGCTGCCT GTCACTGGGC CCTGAGGGAG TCCCAGGGCC AGGACGAAAG TGTGGACAGC AAGAAATCCA CCAGCCATGA CCTGGTCTCG GAGATGGAGG TGTAGGCTGG AGAGACATTG TGGGTGTGTA TCTTCTTATC TCATTTCACA AGACTGGCTT CAGGCATAGC TGGATCCAGG AGCTCAATGA TGTCTTCATT TTATTCCTTC CTTCATTCAA CAGATATCCA TCATGCACTT GCTATGTGCA AGGCCTTTTT AGGCACTAGA GATATAGCAG TGACCAAAAC AGACACAAAT CCTGCCC

SEQ ID NO:34

190701

Cluster name: C-C chemokine receptor 11

SequenceID: NM_016557

Sequence: CAAGACTGCT CCTCTCTGCC GACTACAACA GATTGGAGCC ATGGCTTTGG AGCAGAACCA GTCAACAGAT TATTATTATG AGGAAAATGA AATGAATGGC ACTTATGACT ACAGTCAATA TGAACTGATC TGTATCAAAG AAGATGTCAG AGAATTTGCA AAAGTTTTCC TCCCTGTATT CCTCACAATA GTTTTCGTCA TTGGACTTGC AGGCAATTCC ATGGTAGTGG CAATTTATGC CTATTACAAG AAACAGAGAA CCAAAACAGA TGTGTACATC CTGAATTTGG CTGTAGCAGA TTTACTCCTT CTATTCACTC TGCCTTTTTG GGCTGTTAAT GCAGTTCATG GGTGGGTTTT AGGGAAAATA ATGTGCAAAA TAACTTCAGC CTTGTACACA CTAAACTTTG TCTCTGGAAT GCAGTTTCTG GCTTGTATCA GCATAGACAG ATATGTGGCA GTAACTAAAG TCCCCAGCCA ATCAGGAGTG GGAAAACCAT GCTGGATCAT CTGTTTCTGT GTCTGGATGG CTGCCATCTT GCTGAGCATA CCCCAGCTGG TTTTTTATAC AGTAAATGAC AATGCTAGGT GCATTCCCAT TTTCCCCCGC TACCTAGGAA CATCAATGAA AGCATTGATT CAAATGCTAG AGATCTGCAT TGGATTTGTA GTACCCTTTC TTATTATGGG GGTGTGCTAC TTTATCACAG CAAGGACACT CATGAAGATG CCAAACATTA AAATATCTCG ACCCCTAAAA GTTCTGCTCA CAGTCGTTAT AGTTTTCATT GTCACTCAAC TGCCTTATAA CATTGTCAAG TTCTGCCGAG CCATAGACAT CATCTACTCC CTGATCACCA GCTGCAACAT GAGCAAACGC ATGGACATCG CCATCCAAGT CACAGAAAGC ATCGCACTCT TTCACAGCTG CCTCAACCCA ATCCTTTATG TTTTTATGGG AGCATCTTTC AAAAACTACG TTATGAAAGT GGCCAAGAAA

TATGGGTCCT GGAGAAGACA GAGACAAAGT GTGGAGGAGT
TTCCTTTTGA TTCTGAGGGT CCTACAGAGC CAACCAGTAC
TTTTAGCATT TAAAGGTAAA ACTGCTCTGC CTTTTGCTTG
GATACATATG AATGATGCTT TCCCCTCAAA TAAAACATCT
GCATTATTCT GAAACTCAAA TCTCAGACGC CGTGGTTGCA
ACTTATAATA AAGAATGGGT TGGGGGAAGG GGGAGAAATA
AAAGCCAAGA AGAGGAAACA AGATAATAAA TGTACAAAAC
ATGAAAATTA AAATGAACAA TATAGGAAAA TAATTGTAAC
AGGCATAAGT GAATAACACT CTGCTGTAAC GAAGAAGAGC
TTTGTGGTGA TAATTTTGTA TCTTGGTTGC AGTGGTGCTT
ATACAAATCT ACACAAGTGA TAAAATGACA CAGAACTATA
TACACACATT GTACCAATTT CAATTTCCTG GTTTTGACAT
TATAGTATAA TTATGTAAGA TGGAACCATT GGGGAAAACT
GGGTGAAGGG TACCCAGGAC CACTCTGTAC CATCTTTGTA
ACTTCCTGTG AATTTATAAT AATTTCAAAA TAAAACAAGT
TAAAAAAAAA CCCACTATGC TATAAGTTAG GCCATCTAAA
ACAGATTATT AAAGAGGTTC ATGTTAAAAG GCATTTATAA
TTATTTTTAA TTATCTAAGT TTTAATACAA GAACGATTTC
CCTGCATAAT TTTAGTACTT GAATAAGTAT GCAGCAGAAC
TCCAACTATC TTTTTTCCTG TTTTTTTTAA ATTTGTAAGT

SEQ ID NO:35

190705

Cluster name: G-protein coupled receptor SALPR

SequenceID: NM_016568

Sequence: GATTTGGGGA GTTATGCGCC AGTGCCCCAG TGACCGCGGG
ACACGGAGAG GGGAAGTCTG CGTTGTACAT AAGGACCTAG
GGACTCCGAG CTTGGCCTGA GAACCCTTGG ACGCCGAGTG
CTTGCCTTAC GGGCTGCACT CCTCAACTCT GCTCCAAAGC
AGCCGCTGAG CTCAACTCCT GCGTCCAGGG CGTTCGCTGC
GCGCCAGGAC GCGCTTAGTA CCCAGTTCCT GGGCTCTCTC
TTCAGTAGCT GCTTTGAAAG CTCCCACGCA CGTCCCGCAG
GCTAGCCTGG CAACAAAACT GGGGTAAACC GTGTTATCTT
AGGTCTTGTC CCCCAGAACA TGACCTAGAG GTACCTGCGC
ATGCAGATGG CCGATGCAGC CACGATAGCC ACCATGAATA
AGGCAGCAGG CGGGGACAAG CTAGCAGAAC TCTTCAGTCT
GGTCCCGGAC CTTCTGGAGG CGGCCAACAC GAGTGGTAAC
GCGTCGCTGC AGCTTCCGGA CTTGTGGTGG GAGCTGGGGC
TGGAGTTGCC GGACGGCGCG CCGCCAGGAC ATCCCCCGGG
CAGCGGCGGG GCAGAGAGCG CGGACACAGA GGCCCGGGTG
CGGATTCTCA TCAGCGTGGT GTACTGGGTG GTGTGCGCCC
TGGGGTTGGC GGGCAACCTG CTGGTTCTCT ACCTGATGAA
GAGCATGCAG GGCTGGCGCA AGTCCTCTAT CAACCTCTTC
GTCACCAACC TGGCGCTGAC GGACTTTCAG TTTGTGCTCA
CCCTGCCCTT CTGGGCGGTG GAGAACGCTC TTGACTTCAA
ATGGCCCTTC GGCAAGGCCA TGTGTAAGAT CGTGTCCATG
GTGACGTCCA TGAACATGTA CGCCAGCGTG TTCTTCCTCA
CTGCCATGAG TGTGACGCGC TACCATTCGG TGGCCTCGGC
TCTGAAGAGC CACCGGACCC GAGGACACGG CCGGGGCGAC
TGCTGCGGCC GGAGCCTGGG GGACAGCTGC TGCTTCTCGG
CCAAGGCGCT GTGTGTGTGG ATCTGGGCTT TGGCCGCGCT
GGCCTCGCTG CCCAGTGCCA TTTTCTCCAC CACGGTCAAG
GTGATGGGCG AGGAGCTGTG CCTGGTGCGT TTCCCGGACA
AGTTGCTGGG CCGCGACAGG CAGTTCTGGC TGGGCCTCTA
CCACTCGCAG AAGGTGCTGT TGGGCTTCGT GCTGCCGCTG
GGCATCATTA TCTTGTGCTA CCTGCTGCTG GTGCGCTTCA

TCGCCGACCG CCGCGCGGCG GGGACCAAAG GAGGGGCCGC GGTAGCCGGA GGACGCCCGA CCGGAGCCAG CGCCCGGAGA CTGTCGAAGG TCACCAAATC AGTGACCATC GTTGTCCTGT CCTTCTTCCT GTGTTGGCTG CCCAACCAGG CGCTCACCAC CTGGAGCATC CTCATCAAGT TCAACGCGGT GCCCTTCAGC CAGGAGTATT TCCTGTGCCA GGTATACGCG TTCCCTGTGA GCGTGTGCCT AGCGCACTCC AACAGCTGCC TCAACCCCGT CCTCTACTGC CTCGTGCGCC GCGAGTTCCG CAAGGCGCTC AAGAGCCTGC TGTGGCGCAT CGCGTCTCCT TCGATCACCA GCATGCGCCC CTTCACCGCC ACTACCAAGC CGGAGCACGA GGATCAGGGG CTGCAGGCCC CGGCGCCGCC CCACGCGGCC GCGGAGCCGG ACCTGCTCTA CTACCCACCT GGCGTCGTGG TCTACAGCGG GGGGCGCTAC GACCTGCTGC CCAGCAGCTC

SEQ ID NO:36

190711

Cluster name: G protein-coupled receptor GPR85

SequenceID: NM_018970

Sequence: GGCACGAGGA TTTTACTGCT GTCTCAAGAT CAGATTATTA CTGTAGAGAA GATTTTTATT TTTTGTTTCA TTAACAGATT ATTATAAAGC AAAAAGCATG CAGAAAAAGA AGCAGACGTT TTACATTGGG AATTAATGAA AGCGTGTCTG CTAGTTTTGG GTAGGAGAAC TGGGAAGTTG TTGCTTAAAA TTTTATATCA CCTCCACAAA CAAAACTCTT CGGAAATGGT AAAATAAGAA AATGCATGAT TCTAGAGGCA TTCCTAAGCA CCCACGTGTC AGGCTTTGTG GTGTCTGTGG TATCATCCGA CCGTTTGGAC TGGTTAGGGC TTACTGAGAG CTCCATTTCT GGAAAGCCTT ACAAGACTGA GGAATATCAG ACTGCGAATC ACCGGGAACG GTTCCTTTGC AGCACAGAAG CAATCTCTCT CCCCATCTTC GCATATTCTG ATGGCAAAAC AAGTGGAAGA AAAGAGGAAG CATGACTGCA GATCAGATCA GTTCTCTTTG TGGATTATAT TTTCAGTAAA ATGTATGGAT CTATCTTTTC CTTGTTCTTA TATCTAGATC ATGAGACTTG ACTGAGGCTG TATCCTTATC CTCCATCCAT CTATGGCGAA CTATAGCCAT GCAGCTGACA ACATTTTGCA AAATCTCTCG CCTCTAACAG CCTTTCTGAA ACTGACTTCC TTGGGTTTCA TAATAGGAGT CAGCGTGGTG GGCAACCTCC TGATCTCCAT TTTGCTAGTG AAAGATAAGA CCTTGCATAG AGCACCTTAC TACTTCCTGT TGGATCTTTG CTGTTCAGAT ATCCTCAGAT CTGCAATTTG TTTCCCATTT GTGTTCAACT CTGTCAAAAA TGGCTCTACC TGGACTTATG GGACTCTGAC TTGCAAAGTG ATTGCCTTTC TGGGGGTTTT GTCCTGTTTC CACACTGCTT TCATGCTCTT CTGCATCAGT GTCACCAGAT ACTTAGCTAT CGCCCATCAC CGCTTCTATA CAAAGAGGCT GACCTTTTGG ACGTGTCTGG CTGTGATCTG TATGGTGTGG ACTCTGTCTG TGGCCATGGC ATTTCCCCCG GTTTTAGACG TGGGCACTTA CTCATTCATT AGGGAGGAAG ATCAATGCAC CTTCCAACAC CGCTCCTTCA GGGCTAATGA TTCCTTAGGA TTTATGCTGC TTCTTGCTCT CATCCTCCTA GCCACACAGC TTGTCTACCT CAAGCTGATA TTTTTCGTCC ACGATCGAAG AAAAATGAAG CCAGTCCAGT TTGTAGCAGC AGTCAGCCAG AACTGGACTT TTCATGGTCC TGGAGCCAGT GGCCAGGCAG CTGCCAATTG GCTAGCAGGA TTTGGAAGGG GTCCCACACC ACCCACCTTG CTGGGCATCA GGCAAAATGC AAACACCACA GGCAGAAGAA GGCTATTGGT CTTAGACGAG TTCAAAATGG AGAAAAGAAT CAGCAGAATG TTCTATATAA TGACTTTTCT GTTTCTAACC TTGTGGGGCC CCTACCTGGT GGCCTGTTAT TGGAGAGTTT TTGCAAGAGG GCCTGTAGTA

CCAGGGGGAT TTCTAACAGC TGCTGTCTGG ATGAGTTTTG
CCCAAGCAGG AATCAATCCT TTTGTCTGCA TTTTCTCAAA
CAGGGAGCTG AGGCGCTGTT TCAGCACAAC CCTTCTTTAC
TGCAGAAAAT CCAGGTTACC AAGGGAACCT TACTGTGTTA
TATGAGGGAG CATCTGTAAA TCTTTAGCCT TGTGAAAACT
AACCTTCTCT GCTGAGCAAT TGTGGCCCAT AGCCATATTT
TGAGAAGAAA TTCAAGAATG GAATCAGCAG TTTTAAGGAT
TTGGGCAACA TTCTGCAGTC TTTGCAATAG TTCACCTATA
ATCCTATTTT AAATCTCAGA GTGATCCTGC TGACTGCCAG
CAAAGGTTTG TAATTAAGAA GGGACTGAAC CACTGCCCTA
AGTTTCTTTA TGTGGTCAAA AACTAGATAA TGAAAGTAGC
AGGTGCTAAG TATCAGTGCT AAATGCTCTG TATGTCACTA
CATATGAAAA AACATCAAAA AACAATTAGC ATTGGACATC
TTAATAAATT AAGTTGACAT GAGGTAAATG TGTTGATAAA
AACTAATTTT AGAAGTTTGA AGACTTTAAA ACATTTCATA
CTACTATTGT TTTGCAAAGA CTAAAATATT TGGGGACTTA
AAGTACTGTA ATCCACTAAA GACGTGCCAA TGAATTATTG
GAATATCACA CTTTAAAAAC CGCCTTGTAA GTTCTGGGGA
GCATTCCAAA GCAGTATATT GGTTCCAATT AGAGTTTACT
TTTTTTGTAT TAATACATTG CTATTTCTAA ATACCACTTT
CCTCATCTAC TAGTAAGATT GCTAGCATTG AACTGTATTA
TGTGGTTTTT GTTGATTTGG TATAAAGTTT TTCCAATTCA
TTTATATTTT ACAAATGCTA GATATTGGTC TGGGAGGCAA
CATTAATGGT ACCAGCCTGT CACAACTGAG CAGTTCTAAT
AATGCAGAAT AAATACATGT TGCCTTAAAG GGTTATCTAG
TATCCTTCAT CTTATTTAGC ACTGGAGCAA ATAGCCAAGG
GAAATCAAAT CAGTAACTGG TCATGGTCAT GCATCTAAAA
GTGCATGGAA GATCATTTAT TACTTTTTCC TTTTTTTCTC
ACATGGTTTG AAACTTAAAG TGCACATCAC TGAAATAATG
AGATTTTCTT CTACGGTGTG CTACCCTTTC TAAACTGTTC
TAAGAAGCAG GCAGTTGATG TATGTTTATA TTTTAAGTCA
GCTGTCAAGG GGAGACCACA GCCTTAGTAT GACATCCTGC
ACAATTTGTG AAGCATTTAT TCTACTGAAG GCACAGTCTT
GTTTATACTT TCTGCACATT CAGTGTATTG GTAATTTAAA
TTATTTCAGT TTTAACTTGT GAAAGCTTAT ATTATGATTT
CTGGTATTTT AGAAATACAT TAGAGTCTGT GAGTCTCATT
CTTTAAGATA CAGATGTGTG AACTTCAATA TAAAGTTGCA
TTTGCCAAAA TTTACCCGTG TAGCCTGTTA ATTTTCTTGA
AATAAGTTTT ACATTTTTGG CACATAACAA CGTTTTTTTT
AATTTGGGAG GCAAGCACAA ACTAGGAAGA CTAGCTTTAT
TATGGTTTTG CTTTTTGATT CTTGTAGCTA CTATATTCCA
GACTGGAAAT GTATGAATGA TAATCAACAT AATGCTGATA
AACTGACATA ATATTATCTG TAAAAGCATT ATTTGGTAGT
TTATTATAAT CATCCCTCTA TTATTCTTAA ATGCCAGTAG
TATTTAGAGA TGTGTACCTG CTTAGTTAAT TGGCTCAGAA
TTTTAATATA AACATCACAC TTTAATTTGG AGCATAGTAC
CATAGAAATT TGGGGTTCTA AATATACAAC TTGTAAGAAG
AATGGTTTAC ACTAACATTA TGACAAAACT AGAAAAAGTT
ATTATTTTTG TTTGCTTTCT GTTGTTTTGT TTATTGGTTG
GTTTTTGTGA AGTTTATTTT TTTTTTGGTA TTTGATAATT
AAGATTAGGA ATCTAATAAC ACAGAATTCC ATATTGCTAT
AGTACTTCTG TAAAGAGAAT ATCAATATAA ATAAGGAAAA
TAAATCAATG AAATGTTTCA ATGGTTAAAA AAAAAAAAAA AAAAA

SEQ ID NO:37

190774

Cluster name: Histamine H4 receptor

SequenceID: NM_021624

Sequence: GAATTGTCTG GCTGGATTAA TTTGCTAATT TGACCTTCTT CATCATTTGA TGTGATGCCA GATACTAATA GCACAATCAA TTTATCACTA AGCACTCGTG TTACTTTAGC ATTTTTTATG TCCTTAGTAG CTTTTGCTAT AATGCTAGGA AATGCTTTGG TCATTTTAGC TTTTGTGGTG GACAAAAACC TTAGACATCG AAGTAGTTAT TTTTTTCTTA ACTTGGCCAT CTCTGACTTC TTTGTGGGTG TGATCTCCAT TCCTTTGTAC ATCCCTCACA CGCTGTTCGA ATGGGATTTT GGAAAGGAAA TCTGTGTATT TTGGCTCACT ACTGACTATC TGTTATGTAC AGCATCTGTA TATAACATTG TCCTCATCAG CTATGATCGA TACCTGTCAG TCTCAAATGC TGTGTCTTAT AGAACTCAAC ATACTGGGGT CTTGAAGATT GTTACTCTGA TGGTGGTCGT TTGGGTGCTG GCCTTCTTAG TGAATGGGCC AATGATTCTA GTTTCAGAGT CTTGGAAGGA TGAAGGTAGT GAATGTGAAC CTGGATTTTT TTCGGAATGG TACATCCTTG CCATCACATC ATTCTTGGAA TTCGTGATCC CAGTCATCTT AGTCGCTTAT TTCAACATGA ATATTTATTG GAGCCTGTGG AAGCGTGATC GTCTCAGTAG GTGCCAAAGC CATCCTGGAC TGACTGCTGT CTCTTCCAAC ATCTGTGGAC ACTCATTCAG AGGTAGACTA TCTTCAAGGA GATCTCTTTC TGCATCGACA GAAGTTCCTG CATCCTTTCA TTCAGAGAGA CGGAGGAGAA AGAGTAGTCT CATGTTTTCC TCAAGAACCA AGATGAATAG CAATACAATT GCTTCCAAAA TGGGTTCCTT CTCCCAATCA GATTCTGTAG CTCTTCACCA AAGGGAACAT GTTGAACTGC TTAGAGCCAG GAGATTAGCC AAGTCACTGG CCATTCTCTT AGGGGTTTTT GCTGTTTGCT GGGCTCCATA TTCTCTGTTC ACAATTGTCC TTTCATTTTA TTCCTCAGCA ACAGGTCCTA AATCAGTTTG GTATAGAATT GCATTTTGGC TTCAGTGGTT CAATTCCTTT GTCAATCCTC TTTTGTATCC ATTGTGTCAC AAGCGCTTTC AAAAGGCTTT CTTGAAAATA TTTTGTATAA AAAAGCAACC TCTACCATCA CAACACAGTC GGTCAGTATC TTCTTAAAGA CAATTTTCTC ACCTCTGTAA ATTTTAGTCT CAATC

SEQ ID NO:38

191168

Cluster name: P2Y12 platelet ADP receptor

SequenceID: NM_022788

Sequence: GGCTGCAATA ACTACTACTT ACTGGATACA TTCAAACCCT CCAGAATCAA CAGTTATCAG GTAACCAACA AGAAATGCAA GCCGTCGACA ACCTCACCTC TGCGCCTGGG AACACCAGTC TGTGCACCAG AGACTACAAA ATCACCCAGG TCCTCTTCCC ACTGCTCTAC ACTGTCCTGT TTTTTGTTGG ACTTATCACA AATGGCCTGG CGATGAGGAT TTTCTTTCAA ATCCGGAGTA AATCAAACTT TATTATTTTT CTTAAGAACA CAGTCATTTC TGATCTTCTC ATGATTCTGA CTTTTCCATT CAAAATTCTT AGTGATGCCA AACTGGGAAC AGGACCACTG AGAACTTTTG TGTGTCAAGT TACCTCCGTC ATATTTTATT TCACAATGTA TATCAGTATT TCATTCCTGG GACTGATAAC TATCGATCGC TACCAGAAGA CCACCAGGCC ATTTAAAACA TCCAACCCCA AAAATCTCTT GGGGGCTAAG ATTCTCTCTG TTGTCATCTG GGCATTCATG TTCTTACTCT CTTTGCCTAA CATGATTCTG ACCAACAGGC AGCCGAGAGA CAAGAATGTG AAGAAATGCT CTTTCCTTAA ATCAGAGTTC GGTCTAGTCT GGCATGAAAT AGTAAATTAC ATCTGTCAAG TCATTTTCTG GATTAATTTC TTAATTGTTA TTGTATGTTA TACACTCATT ACAAAAGAAC TGTACCGGTC ATACGTAAGA ACGAGGGGTG TAGGTAAAGT CCCCAGGAAA AAGGTGAACG TCAAAGTTTT CATTATCATT GCTGTATTCT TTATTTGTTT TGTTCCTTTC CATTTTGCCC

GAATTCCTTA CACCCTGAGC CAAACCCGGG ATGTCTTTGA CTGCACTGCT GAAAATACTC TGTTCTATGT GAAAGAGAGC ACTCTGTGGT TAACTTCCTT AAATGCATGC CTGGATCCGT TCATCTATTT TTTCCTTTGC AAGTCCTTCA GAAATTCCTT GATAAGTATG CTGAAGTGCC CCAATTCTGC AACATCTCTG TCCCAGGACA ATAGGAAAAA AGAACAGGAT GGTGGTGACC CAAATGAAGA GACTCCAATG TAAACAAATT AACTAAGGAA ATATTTCAAT CTCTTTGTGT TCAGAACTCG TTAAAGCAAA GCGCTAAGTA AAAATATTAA CTGACGAAGA AGCAACTAAG TTAATAATAA TGACTCTAAA GAAACAGAAG ATTACAAAAG CAATTTTCAT TTACCTTTCC AGTATGAAAA GCTATCTTAA AATATAGAAA ACTAATCTAA ACTGTAGCTG TATTAGCAGC AAAACAAACG AC

SEQ ID NO:39

191218

Cluster name: G protein-coupled receptor Ls191218

SequenceID: AX099247

Sequence: TTAATCTCTT CAAGCCTCTG ATTTCCTCTC CTGTAAAACA GGGGCGGTAA TTACCACATA ACAGGCTGGT CATGAAAATC AGTGAACATG CAGCAGGTGC TCAAGTCTTG TTTTTGTTTC CAGGGGCACC AGTGGAGGTT TTCTGAGCAT GGATCCAACC ACCCCGGCCT GGGGAACAGA AAGTACAACA GTGAATGGAA ATGACCAAGC CCTTCTTCTG CTTTGTGGCA AGGAGACCCT GATCCCGGTC TTCCTGATCC TTTTCATTGC CCTGGTCGGG CTGGTAGGAA ACGGGTTTGT GCTCTGGCTC CTGGGCTTCC GCATGCGCAG GAACGCCTTC TCTGTCTACG TCCTCAGCCT GGCCGGGGCC GACTTCCTCT TCCTCTGCTT CCAGATTATA AATTGCCTGG TGTACCTCAG TAACTTCTTC TGTTCCATCT CCATCAATTT CCCTAGCTTC TTCACCACTG TGATGACCTG TGCCTACCTT GCAGGCCTGA GCATGCTGAG CACCGTCAGC ACCGAGCGCT GCCTGTCCGT CCTGTGGCCC ATCTGGTATC GCTGCCGCCG CCCCAGACAC CTGTCAGCGG TCGTGTGTGT CCTGCTCTGG GCCCTGTCCC TACTGCTGAG CATCTTGGAA GGGAAGTTCT GTGGCTTCTT ATTTAGTGAT GGTGACTCTG GTTGGTGTCA GACATTTGAT TTCATCACTG CAGCGTGGCT GATTTTTTTA TTCATGGTTC TCTGTGGGTC CAGTCTGGCC CTGCTGGTCA GGATCCTCTG TGGCTCCAGG GGTCTGCCAC TGACCAGGCT GTACCTGACC ATCCTGCTCA CAGTGCTGGT GTTCCTCCTC TGCGGCCTGC CCTTTGGCAT TCAGTGGTTC CTAATATTAT GGATCTGGAA GGATTCTGAT GTCTTATTTT GTCATATTCA TCCAGTTTCA GTTGTCCTGT CATCTCTTAA CAGCAGTGCC AACCCCATCA TTTACTTCTT CGTGGGCTCT TTTAGGAAGC AGTGGCGGCT GCAGCAGCCG ATCCTCAAGC TGGCTCTCCA GAGGGCTCTG CAGGACATTG CTGAGGTGGA TCACAGTGAA GGATGCTTCC GTCAGGGCAC CCCGGAGATG TCGAGAAGCA GTCTGGTGTA GAGATGGACA GCCTCTACTT CCATCAGATA TATGTG

SEQ ID NO:40

189884

Cluster name: G protein-coupled receptor LS189884

SequenceID: ENSMDNA108574

Sequence: ATGCTGGCAG CTGCCTTTGC AGACTCTAAC TCCAGCAGCA TGAATGTGTC CTTTGCTCAC CTCCACTTTG CCGGAGGGTA CCTGCCCTCT GATTCCCAGG ACTGGAGAAC

CATCATCCCG GCTCTCTTGG TGGCTGTCTG CCTGGTGGGC TTCGTGGGAA ACCTGTGTGT GATTGGCATC CTCCTTCACA ATGCTTGGAA AGGAAAGCCA TCCATGATCC ACTCCCTGAT TCTGAATCTC AGCCTGGCTG ATCTCTCCCT CCTGCTGTTT TCTGCACCTA TCCGAGCTAC GGCGTACTCC AAAAGTGTTT GGGATCTAGG CTGGTTTGTC TGCAAGTCCT CTGACTGGTT TATCCACACA TGCATGGCAG CCAAGAGCCT GACAATCGTT GTGGTGGCCA AAGTATGCTT CATGTATGCA AGTGACCCAG CCAAGCAAGT GAGTATCCAC AACTACACCA TCTGGTCAGT GCTGGTGGCC ATCTGGACTG TGGCTAGCCT GTTACCCCTG CCGGAATGGT TCTTTAGCAC CATCAGGCAT CATGAAGGTG TGGAAATGTG CCTCGTGGAT GTACCAGCTG TGGCTGAAGA GTTTATGTCG ATGTTTGGTA AGCTCTACCC ACTCCTGGCA TTTGGCCTTC CATTATTTTT TGCCAGCTTT TATTTCTGGA GAGCTTATGA CCAATGTAAA AAACGAGGAA CTAAGACTCA AAATCTTAGA AACCAGATAC GCTCAAAGCA AGTCACAGTG ATGCTGCTGA GCATTGCCAT CATCTCTGCT CTCTTGTGGC TCCCCGAATG GGTAGCTTGG CTGTGGGTAT GGCATCTGAA GGCTGCAGGC CCGGCCCCAC CACAAGGTTT CATAGCCCTG TCTCAAGTCT TGATGTTTTC CATCTCTTCA GCAAATCCTC TCATTTTTCT TGTGATGTCG GAAGAGTTCA GGGAAGGCTT GAAAGGTGTA TGGAAATGGA TGATAACCAA AAAACCTCCA ACTGTCTCAG AGTCTCAGGA AACACCAGCT GGCAACTCAG AGGGTCTTCC TGACAAGGTT CCATCTCCAG AATCCCCAGC ATCCATACCA GAAAAAGAGA AACCCAGCTC TCCCTCCTCT GGCAAAGGGA AAACTGAGAA GGCAGAGATT CCCATCCTTC CTGACGTAGA GCAGTTTTGG CATGAGAGGG ACACAGTCCC TTCTGTACAG GACAATGACC CTATCCCCTG GGAACATGAA GATCAAGAGA CAGGGGAAGG TGTTAAATAG

SEQ ID NO:41

168928

Cluster name: G protein-coupled receptor Ls168928

SequenceID: AW973537

Sequence: AGTAGTAATC TCATCTTGTG CACTGTGGGG TCTTCTAATG TGACCCTGAG CAATCTTCTG CATACCAGTA AAGACTGTTC ACTTTTCCAC CATGAACTCC ATCATCAGAA GACTGTTTCT TACTCTGTTT CTTACTCCAG ATATGTTTTT CTTATAGGAA CAATGCTGCT TTCAAGTGCA TACAGAGTGG TCCTTTTGTT CAGGCACCAG AAGAAATTCT GATACTTTCA CAGCACCAGC CTTTCCCCAA GACCTTCCCC AGAGAAAAGT GCCACTCAGA CCATCCTGCT GCTAGTGAGT TTCTTTGTGG TCATCTACTG GGTCGATTTC ATCATCTCAT GCACCTCAAC CTTGCTATGG GCATATGACC CTGTTGTCCT GGGTGTCCAG AGGCTTGTCA GTCTTTTGGT GCTACTCAGA TCTGATAAAA GGATAATCAT TGTGACACAA ACTGTGAGAC AGATGGTTAA CAAGTTATTT TTATTGAAAA TAGATTATTC TGTCACCAGT TAAATTACAT AAGTAGTACA GAACTTGCTA TTTAATTAAC TTAAATGGTT GGATTTACAC TTTCAATATG

SEQ ID NO:42

189890

Cluster name: G protein-coupled receptor Ls189890

SequenceID: ENSMDNA279706

Sequence: CTTCCTCATC AGACTGTTGC CTGGCTACAC GGCTGGGCGC AGCGCCAACA GGAAGTCCTT AAAGGCAGGT ATTATTCCTA AGTGTATGGT CAGGCTCAAG CTGCCATTCA GCAACTCGTG GGCTTTGGGA CCCAGCACCG AGGGGTTATA TGTGAAGGAG GGCCCCCGCC AGGAGTCTGA AGTGAAAATG GTAGCAGTCA CAGACAATGA CGGTGGCAGC AGGGGTTTAG GCAATGACGG TGGCCATGCT GTTGATGCTG TCATCTACAC TGCTGATCTT TGA

SEQ ID NO:43

189893

Cluster name: G protein-coupled receptor Ls189893

SequenceID: AI285887

Sequence: TTTGTGTACA AGAATTTTAT GTACTTTAAC TACTGTGGCA CAAGTGACAT GGCCAAAATG GACCTTTCCT CCAACACACT GGTGCTGTGG CGTCTGCTGC CTGGTGCCAC CTATAACAAC CGCTTTTCCT ATGCTGGTGT GCCCTGGAAG GACTTAGATT TTGCTGGTGA TGAGAAGGGG CTGTGGGTTC TCTATGCCAC TGAGGAGAGC AAGGGCAACC TGGTTGTGAG TCGTCTCAAC GCTAGCACCC TAGAAGTGGA GAAAACCTGG CGTACCAGCC AGTACAAGCC AGCCCTGTCA GGGGCCTTCA TGGCCTGTGG GGTGCTCTAT GCCTTACACT CACTGAACAC CCACCAAGAG GAGATCTTCT ATGCTTTTGA CACCACCACC GGG

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/15332

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C07K 14/705, 16/28; C12N 15/12
US CL : 435/7.1, 69.1, 252.3, 320.1; 530/350, 388.22; 536/23.5
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/7.1, 69.1, 252.3, 320.1; 530/350, 388.22; 536/23.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Geneseq, Issued Patents, EST
searched SEQ ID NO:3

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A,P	WHITE et al. (The ADHR Consortium), Autosomal dominant hypophosphataemic rickets is associated with mutations in FGF23. Nature Genetics. November 2000. Vol. 26. No.3. pages 345-348. see entire document.	1-10, 14-18
A,P	WO 01/04292 A1 (MERCK PATENT GMBH) 18 January 2001. SEQ ID NO:1.	1-10, 14-18

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 SEPTEMBER 2001	25 OCT 2001

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
JOHN ULM
Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

Inte[r]nal application No.
PCT/US01/15332

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-10 and 14 to 18 in so far as they relate to SEQ ID NO:3.

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack Unity of Invention because they are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for more than one species to be searched, the appropriate additional search fees must be paid. The different species consist of the 48 nucleotide sequences listed in Table 1 of the instant description and 48 antibodies which bind to 48 *different polypeptides*.

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I-XLVIII, claims 1 to 10 and 14 to 18, which are drawn to an isolated polynucleotide encoding any one of 48 different polypeptides, an isolated polypeptide encoded by that nucleic acid and methods of use.
Group II, XLIX-XCVI, claims 11 to 13, drawn to an antibody which binds to any one of 48 different polypeptides.

The inventions listed as Groups I and II do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the nucleic acids and proteins of invention I do not share a common utility with the antibodies of invention II and each of these inventions can be made and used without the other.

The species listed above do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons: The 48 nucleic acids listed in Table 1 of the instant description lack a common utility which is based upon a special technical feature which is common to all of those nucleic acids and which is lacking from the prior art.